–1 9 9 4 6

SEARCH REQUEST FORM

Examiner # (Mandatory): _____   Requester's Full Name: _____

Art Unit _____ Location (Bldg/Room#): _____ 8817 _____   Phone (circle 305 306 308) _____

Serial Number: _____   Results Format Preferred (circle): PAPER DISK E-MAIL

Title of Invention _____

Inventors (please provide full names): _____

_____

Earliest Priority Date: _____

Keywords (include any known synonyms registry numbers, explanation of initialisms):

Search Topic:

Please write detailed statement of the search topic, and the concept of the invention. Describe as specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc., if known. You may include a copy of the abstract and the broadcast or most relevant claim(s).

STAFF USE ONLY

Searcher: _____	**Type of Search**	**Vendors** (include cost where applicable)
Searcher Phone #: _____	_4_ N.A. Sequence	_____ STN
Searcher Location: _____	_1_ A.A. Sequence	_____ Questel/Orbit
Date Picked Up: _9/24_	_____ Structure (#)	_____ Lexis/Nexis
Date Completed: _9/29_	_____ Bibliographic	_____ WWW/Internet
Clerical Prep Time: _24_	_____ Litigation1	_✓_ In-house sequence systems (list)
Terminal Time: _23_	_____ Fulltext	_____ Dialog
Number of Databases: _11_	_____ Procurement	_____ Dr. Link
	_____ Other	_____ Westlaw *abssol*
		_____ Other (specify) *mp2*

ABSS02 FYLER 606

This Page Blank (uspto)

```
**************************************************
*                                                *
*   |\/| |¯¯| |¯¯¯ |¯¯¯ |¯¯| |¯¯|                 *
*   |  | |  | |__  |__  |  | |__                  *
*                                     (TM)        *
**************************************************
```

```
        Release 3.1A John F. Collins, Biocomputing Research Unit.
        Copyright (c) 1993-1998 University of Edinburgh, U.K.
                Distribution rights by Oxford Molecular Ltd
```

```
MPsrch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
           which has been backtranslated into n.a. using IUPAC symbols

Run on:           Wed Sep 29 02:16:56 1999;  MasPar time 33.26 Seconds
                                             1016.259 Million cell updates/sec

Tabular output not generated.

Title:          >US-09-030-606-172
Description:    (1-159) from US09030606.pep
Perfect Score:  1871
N.A. Sequence:     1 ATGGTNGARGCNWSNYTNWS.........ARARARACNGTNCARGCNWSN 477
        Comp:        TACCANCTYCGNWSNRAWS.........TYTYTGNCANGTYCGNWSN

Scoring table:  TABLE bktranslate2
                Gap 30

Nmatch STD :    Dbase 0; Query 0

Searched:       137068 seqs, 35432894 bases x 2

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:       n-issued
                1:5A_COMB 2:5B_COMB 3:5C_COMB 4:PCT9_COMB 5:backfiles1

Statistics:     Mean 45.107;  Variance 172.683;  scale 0.261

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
	1	1769	94.5	871	3	US-08-744-	Sequence 2, Applicatio	1.34e-174
	2	489	26.1	732	3	US-08-361-	Sequence 2, Applicatio	2.20e-35
c	3	477	25.5	1089	4	PCT-US96-0	Sequence 1, Applicatio	3.87e-34
c	4	477	25.5	1089	4	PCT-US96-0	Sequence 3, Applicatio	3.87e-34
	5	316	16.9	699	3	US-08-738-	Sequence 8, Applicatio	8.81e-18
	6	286	15.3	957	3	US-08-684-	Sequence 11, Applicati	7.72e-15
	7	266	14.2	988	3	US-08-684-	Sequence 10, Applicati	6.65e-13
	8	263	14.1	992	2	US-08-358-	Sequence 13, Applicati	1.29e-12
	9	263	14.1	1462	3	US-08-358-	Sequence 14, Applicati	1.29e-12
	10	255	13.6	840	3	US-08-684-	Sequence 12, Applicati	7.52e-12
	11	253	13.5	760	4	PCT-US95-0	Sequence 7, Applicati	1.17e-11
	12	253	13.5	766	4	PCT-US96-0	Sequence 9, Applicatio	1.17e-11
	13	253	13.5	832	4	PCT-US95-0	Sequence 8, Applicatio	1.17e-11
	14	250	13.4	1096	3	US-08-684-	Sequence 9, Applicatio	2.25e-11
	15	225	12.0	1333	3	US-08-684-	Sequence 9, Applicatio	5.06e-09
	16	219	11.7	1454	4	US-08-467-	Sequence 2, Applicatio	1.82e-08
	17	219	11.7	1454	4	PCT-US96-0	Sequence 2, Applicatio	1.82e-08
	18	215	11.5	734	3	US-08-650-	Sequence 1, Applicatio	4.26e-08
	19	215	11.5	821	3	US-08-650-	Sequence 2, Applicatio	4.26e-08
	20	215	11.5	866	3	US-08-650-	Sequence 3, Applicatio	4.26e-08
	21	195	10.4	357	2	US-08-487-	Sequence 1, Applicatio	2.83e-06
c	22	191	10.2	7218	2	US-08-232-	Sequence 14, Applicati	6.47e-06
	23	179	9.6	771	4	PCT-US95-0	Sequence 22, Applicati	7.51e-05
	24	179	9.6	771	4	PCT-US95-0	Sequence 3, Applicatio	7.51e-05
	25	178	9.5	779	4	PCT-US95-1	Sequence 80, Applicati	9.20e-05
	26	173	9.2	681	4	US-07-929-	Sequence 3, Applicatio	2.52e-04
	27	172	9.2	841	4	PCT-US95-1	Sequence 110, Applicat	3.08e-04
	28	169	9.0	797	5	5180819-1	Patent No. 5180819.	5.60e-04
	29	169	9.0	835	5	5223425-7	Patent No. 5223425.	5.60e-04
	30	169	9.0	894	4	PCT-US95-1	Sequence 90, Applicati	5.60e-04
	31	169	9.0	1093	5	5223425-3	Patent No. 5223425.	5.60e-04
	32	164	8.8	1233	3	US-08-286-	Sequence 1, Applicatio	1.51e-03
	33	164	8.8	1233	1	US-08-254-	Sequence 7, Applicatio	1.51e-03
	34	164	8.8	1236	1	US-07-957-	Sequence 17, Applicati	1.51e-03
	35	164	8.8	1236	3	US-08-153-	Sequence 2, Applicatio	1.51e-03
	36	164	8.8	1335	2	US-07-942-	Sequence 2, Applicatio	1.51e-03
	37	164	8.8	1372	5	5219569-1	Patent No. 5219569.	1.51e-03
	38	164	8.8	2301	5	5188829-2	Patent No. 5188829.	1.51e-03
	39	163	8.7	207	2	US-08-485-	Sequence 76, Applicati	1.84e-03
	40	163	8.7	207	4	PCT-US95-1	Sequence 76, Applicati	1.84e-03
	41	162	8.7	225	4	PCT-US95-1	Sequence 100, Applicati	1.84e-03
	42	163	8.7	378	4	PCT-US95-1	Sequence 96, Applicati	2.24e-03
	43	163	8.7	681	1	US-07-929-	Sequence 5, Applicatio	1.84e-03
	44	163	8.7	681	1	US-07-929-	Sequence 1, Applicatio	1.84e-03
	45	158	8.4	2033	2	US-08-448-	Sequence 14, Applicati	4.90e-03

ALIGNMENTS

```
RESULT  1
ID   US-08-744-026-2 STANDARD; DNA; UNC; 871 BP.
AC   xxxxxx
DT
DE   Sequence 2, Application US/08744026
CC   Sequence 2, Application US/08744026
CC   Patent No. 5786148
CC   GENERAL INFORMATION:
CC     APPLICANT: Bandman, Olga
CC     APPLICANT: Goli, Surya K.
CC     TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
CC     TITLE OF INVENTION: KALLIKREIN
CC     NUMBER OF SEQUENCES: 5
CC     CORRESPONDENCE ADDRESS:
CC       ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
CC       STREET: 3174 Porter Drive
CC       CITY: Palo Alto
CC       STATE: CA
CC       COUNTRY: US
CC       ZIP: 94304
CC     COMPUTER READABLE FORM:
CC       MEDIUM TYPE: Diskette
CC       COMPUTER: IBM Compatible
CC       OPERATING SYSTEM: DOS
CC       SOFTWARE: FastSEQ Version 2.0
CC     CURRENT APPLICATION DATA:
CC       APPLICATION NUMBER: US/08/744,026
CC       FILING DATE: Herewith
CC       CLASSIFICATION: 514
CC     PRIOR APPLICATION DATA:
CC       APPLICATION NUMBER:
CC       FILING DATE:
CC     ATTORNEY/AGENT INFORMATION:
CC       NAME: Billings, Lucy J
CC       REGISTRATION NUMBER: 36,749
CC       REFERENCE/DOCKET NUMBER: PF-0154 US
CC     TELECOMMUNICATION INFORMATION:
CC       TELEPHONE: 415-855-0555
CC       TELEFAX: 415-845-4166
CC   INFORMATION FOR SEQ ID NO: 2:
CC     SEQUENCE CHARACTERISTICS:
CC       LENGTH: 871 base pairs
CC       TYPE: nucleic acid
```

```
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   IMMEDIATE SOURCE:
CC   LIBRARY:
CC   CLONE: Consensus
SQ   SEQUENCE 871 BP; 166 A; 260 C; 258 G; 184 T; 3 OTHER.

Query Match        94.5%;   Score 1769;   DB 3;   Length 871;
Best Local Similarity   55.3%;   Pred. No. 1.34e-174;
Matches   263;   Conservative  113;   Mismatches 100;   Indels 0;   Gaps 0;

Db  298  ATGGTGGAGGCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTCGCTAAC  357
         |||||  ||||  |||   ::  :|   :| ::| |:||:||||::|| ||  ||| ||
Qy    1  ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY  60

Db  358  GACCTCATGNTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACAACATCCGAGNATC  417
         || ||| |||  ||   :| | |:||:|| ::|:||   | |:|| || ||  ::||||
Qy   61  GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH  120

Db  418  AGCATTGNTTCCCAGTGCCCTACCGCGGGGAACTTTGCCTCGTTCTGGGCTGGGTCTG  477
         ::||| ||||  | ::|||    ||  | |||  | |||   :| || |  |||| :|
Qy  121  WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGNTGGGGNYTN  180

Db  478  CTGGCGAACGGCAGAAGTGCCTACCGTGCTGCAGTGCGTGTGGTGTGTCTGAG  537
         :| |||| | ||| |:|| || ||  |:| |:| |:||  | :|| |:||
Qy  181  YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNWSNGAR  240

Db  538  GAGGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATGTTCTGCGCCGCGGA  597
         | ||| ||| :|  :| |: |||||  | | ::|| ::| ::| |: :|:|||| ||
Qy  241  GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGN  300

Db  598  GGGCAAGACCAGAAGGACTCCTGCAACGGGTGACTCTGGGGCGCCCCTGACTCTGCAACGGG  657
         ||| |:|| ||| || ||: || |::| |:|:| :|:|  |||| :|   |:||| |||
Qy  301  GGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGN  360

Db  658  TACTTGCAGGGCCTTGTGTCTTTCGGAAAAGCCCCGTGTGGCCGTGCCAGGT  717
         |||||||| ||: ||  :||  | || |: ::|:|:|||||  ||:|| |||
Qy  361  TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGN  420

Db  718  GTCTACCACCAACCTCTGCAAATTCACTGAGTGGATGAGAAAACCGTCCAGGCCAG  773
         ||:||:||:||  |  |::||  ||:|| |:|:|||  |: |:|| || |||||
Qy  421  GTNTAYCACNAAYYTNTGYAARTTYACNGARTGGATGARARAARACNGTNCARGCNWS  476
```

```
CC   CLASSIFICATION: 530
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Blalock, Donna K.
CC   REGISTRATION NUMBER: 38,082
CC   REFERENCE/DOCKET NUMBER: X8350A
CC   TELECOMUNICATION INFORMATION:
CC   TELEPHONE: 317-277-1090
CC   TELEFAX: 317-276-3861
CC   INFORMATION FOR SEQ ID NO: 2:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 732 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: both
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA
SQ   SEQUENCE 732 BP; 168 A; 212 C; 203 G; 149 T; 0 OTHER.

Query Match        26.1%;   Score 489;   DB 3;   Length 732;
Best Local Similarity   36.2%;   Pred. No. 2.20e-35;
Matches   165;   Conservative  71;   Mismatches 220;   Indels 0;   Gaps 0;

Db  275  CTGTGATCCACCCTGACTATGATGCCGCCAGCCATGACCAGGACATCATGCTGTTGCGCC  334
         : |: || |||| :| || |||  ||| ||  |: | |  || ||:|| | ||:||
Qy   20  SNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARY  79

Db  335  TGGCACGCCCAGCCAAACTCTGAACTCATCCAGCCCCTTCCCGGAGAGGGACTGCT  394
         | |: |||  |||| ||:| || |: || ||| |:|    ::    |:|| :|:||
Qy   80  TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHGCNWSNCARTGYC  139

Db  395  CAGCCAACACCACCAGCTGCCACATCCTGGCTGGGCAAGACAGCAGATGGTGATTTCC  454
         || || | | |||| : |:||| || || |:| |:|:|| | | |:| |:|| ||::
Qy  140  CNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTNYTNGCNAAYGGNMGNATGC  199

Db  455  CTGACACCATCCAGTGCATACATCCACCTGGTCTCCGTGAGGAGTGTGAGCATGCCT  514
         |:|| ||:||||| || :|| || ||:|  |  || :|:|: |:| | |:|| ||
Qy  200  CNACNGTNYTNCARTGYGTNAAYGTNWSNGTNWSNGARGARTNTGYWSNAARYTNT  259

Db  515  ACCCTGCCAGATCACCCAGAACATGTTGTGCTGGGATGAGAAGTACGGGAAGGATT  574
         ||:|  | |  | |  | ||  || |||:|| || | :|:|:|| |:|||| ||
Qy  260  AYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGNGGNCARNNNCARNNNGAYW  319

Db  575  CCTGCCAGGTGATTCTGGGGGGTCCGCTGGTGTTGGAGACCACCTCCGAGGCCTTGTGT  634
         ||||| |:|: | ||  |: |:|| :|   |: |:|| || || :||| |||| ||:|
Qy  320  SNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNW  379

Db  635  CATGGGGTAACATCCCCTGTGGATCAAAGAGAAGCCAGGAGTCTACACCAACGTCTGCA  694
         ||||| |||  || |   ||:||||  |: || || || |:||:||:||  |:|| |
Qy  380  SNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNTGYA  439

Db  695  GATACACGAACTGGATCCAAAAAACCATTCAGGCCA  730
         ||:||: | | ||| || |||| | || |||:|||
Qy  440  ARTTYACNGARTGGATHGARARARACNGTNCARGCNW  475
```

```
RESULT 2
ID   US-08-361-395-2 STANDARD; DNA; UNC; 732 BP.
AC   xxxxxxx
DT
DE   Sequence 2, Application US/08361395
CC   Sequence 2, Application US/08361395
CC   Patent No. 5733768
CC   GENERAL INFORMATION:
CC   APPLICANT: Dixon, Eric P.
CC   APPLICANT: Johnstone, Edward M.
CC   APPLICANT: Little, Sheila P.
CC   APPLICANT: No. 5733768ris, Franklin H.
CC   TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS
CC   NUMBER OF SEQUENCES: 8
CC   CORRESPONDENCE ADDRES:
CC   ADDRESSEE: Eli Lilly and Company
CC   STREET: Lilly Corporate Center/Patent Division
CC   CITY: Indianapolis
CC   STATE: Indiana
CC   COUNTRY: United States
CC   ZIP: 46285
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.30
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/361,395
CC   FILING DATE: 22-DEC-1994
```

```
RESULT 3
ID   PCT-US96-04294-1 STANDARD; DNA; UNC; 1089 BP.
AC   xxxxxx
DT
DE   Sequence 1, Application PC/TUS9604294
CC   Sequence 1, Application PC/TUS9604294
CC   GENERAL INFORMATION:
CC   APPLICANT: Dixon, Eric P.
CC   APPLICANT: Johnstone, Edward M.
CC   APPLICANT: Little, Sheila P.
CC   TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
CC   TITLE OF INVENTION: RELATED NUCLEIC ACIDS
CC   NUMBER OF SEQUENCES: 3
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSE: Eli Lilly and Company
CC   STREET: Lilly Corporate Center
CC   CITY: Indianapolis
CC   STATE: Indiana
CC   COUNTRY: United States of America
```

CC   APPLICANT: Dixon, Eric P.
CC   APPLICANT: Johnstone, Edward M.
CC   APPLICANT: Little, Sheila P.
CC   TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
CC   TITLE OF INVENTION: RELATED NUCLEIC ACIDS
CC   NUMBER OF SEQUENCES: 3
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Eli Lilly and Company
CC   STREET: Lilly Corporate Center
CC   CITY: Indianapolis
CC   STATE: Indiana
CC   COUNTRY: United States of America
CC   ZIP: 46285
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.30
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: PCT/US96/04294
CC   FILING DATE:
CC   CLASSIFICATION:
CC   PRIOR APPLICATION DATA:
CC   APPLICATION NUMBER: US 08/416,257
CC   FILING DATE: 04-APR-1995
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Blalock, Donna K.
CC   REGISTRATION NUMBER: 38,082
CC   REFERENCE/DOCKET NUMBER: X9239
CC   TELECOMMUNICATION INFORMATION:
CC   TELEPHONE: 317-277-1090
CC   TELEFAX: 317-276-3861
CC   INFORMATION FOR SEQ ID NO: 3:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 1089 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: mRNA
SQ   SEQUENCE 1089 BP: 218 A; 263 C; 319 G; 0 T; 289 OTHER.

Query Match
Best Local Similarity 25.5%;   Score 477; DB 4; Length 1089;
                       27.4%;   Pred. No. 3.87e-34;
Matches 129; Conservative 128; Mismatches 207; Indels 6; Gaps 4;

Db 224 UUUCAUGGUGUCAUUUAUCCACUUGGUGAACUUGCGAGACUUGAGUCGAGACUCCUGGGUC 283
            :: || | | | ||||| | |||:|| | | |||||
Cp 471 YTGNACNGTYTYTCDATCCAYTCNGTRAAYTTRCANARRTTNGTRTANACNCCNGGNAC 412

Db 284 AUUGGGUUGGCCGCAAGGGAAAGUUCCCCAGGACACCAGACCUUGCAGGGUACCUCUGCA 343
            :|| ||||| || || || || :: || ||||| |::||:|
Cp 411 NCCNACYTGNCCRCANGGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCCRTTRCA 352

Db 344 CACCAACGGUCCCCCUGAGUCACCAUUGCAGGCGUUUUCUUGGAGUCGGGGAUGCCAGC 403
            :|:|| || ||||| :|| :||||:: || |||: | |:| |
Cp 351 DATNARNGGNCCNCCNSWRTCNCCRTTRCANSWRTCNNNYTGNNNYTGNCCNCCNCNGC 292

Db 404 GCACAGCAUGGAAUUUCCAGUAAGCCUUGUAAACCUUCGUCAGUCCUGGGGGGAGAU 463
            :||:|| ||| : |: | :|| :|| | :|| ||| || ::||
Cp 291 RCARAACATNSWNGGRTGRTANARNGGRTCRTANARYTTNSWRCANACYTCYTCNSWNAC 232

Db 464 GAGCUUGACAUCCACCCACAUGAGGUCAGAGGGAAAGUCACAUCUGGGCUCGUGUGUAGU 523
            : || || :| || |: :| || |: | || || |: |: :| |
Cp 231 NACNSWNACRTTNACRCAYTGNARNACNGTNGGCCATNCKNCCRT-TNG--CNARN-ARN 177

Db 524 GCCCCAGCCGGAGACAGUACAGGUGGUUCCAGCGGGUUCCAGGGGUUCCNGGGGGAUGCCAGCCUGAC 583
            ||||| ||| ::|| | |||| | |: || ||| ||| : | || |
Cp 176 -CCCCANCCNSWNACNARRCANSWRTTNCCNGCNGTNGGRCAYTGNSWNGCDATNSWDAT 118

Db 584 UUUCUUCACCAUGGAUGACAGCCUGCCUAUUGAGCUUCACGAGCAUGAGGUCAUU 643
            :|||| || :|: : ||::| | |||: |: | |
Cp 117 NSWNCKDATNGTRTCNSWYTCNSWNACNSWYTCRTCNARYTTDATNARCATNARRTCRTT 58

CC   ZIP: 46285
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.30
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: PCT/US96/04294
CC   FILING DATE:
CC   CLASSIFICATION:
CC   PRIOR APPLICATION DATA:
CC   APPLICATION NUMBER: US 08/416,257
CC   FILING DATE: 04-APR-1995
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Blalock, Donna K.
CC   REGISTRATION NUMBER: 38,082
CC   REFERENCE/DOCKET NUMBER: X9239
CC   TELECOMMUNICATION INFORMATION:
CC   TELEPHONE: 317-277-1090
CC   TELEFAX: 317-276-3861
CC   INFORMATION FOR SEQ ID NO: 1:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 1089 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA
SQ   SEQUENCE 1089 BP: 218 A; 263 C; 319 G; 289 T; 0 OTHER.

Query Match
Best Local Similarity 36.8%;   Score 477; DB 4; Length 1089;
                                      Pred. No. 3.87e-34;
Matches 173; Conservative 84; Mismatches 207; Indels 6; Gaps 4;

Db 224 TTTCATGGTGTCATTTATCCACTTGGTGAACTTGCGAGACTTGAGTCGAGACTCCTGGGTC 283
            :| | || | | | ||||| | |||:|| | | ||||| |||
Cp 471 YTGNACNGTYTYTCDATCCAYTCNGTRAAYTTRCANARRTTNGTRTANACNCCNGGNAC 412

Db 284 ATTGGGTTGGCCGCAAGGGAAAGTTCCCCAGGACACCAGACCTTGCAGGGTACCTCTGCA 343
            :|| ||||| || || || || :: || ||||| |:: ||:|
Cp 411 NCCNACYTGNCCRCANGGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCCRTTRCA 352

Db 344 CACCAACGGTCCCCCTGAGTCACCATTGCAGGCGTTTTCTTGGAGTCGGGGATGCCAGC 403
            :|:|| || ||||| :|| :||||:: || |||: | |:| |
Cp 351 DATNARNGGNCCNCCNSWRTCNCCRTTRCANSWRTCNNNYTGNNNYTGNCCNCCNCNGC 292

Db 404 GCACAGCATGGAATTTCCAGTAAGCCTTGTAAACCTTCGTCAGTCCTGGGGGGAGAT 463
            :||:|| ||| : |: | :|| :|| | :|| ||| || ::||
Cp 291 RCARAACATNSWNGGRTGRTANARNGGRTCRTANARYTTNSWRCANACYTCYTCNSWNAC 232

Db 464 GAGCTTGACATCCACCCACATGAGGTCAGAGGGAAAGTCACATCTGGGCTCGTGTGTAGT 523
            : || || :| || |: :| || |: | || || |: |: :| |
Cp 231 NACNSWNACRTTNACRCAYTGNARNACNGTNGGCCATNCKNCCRT-TNG--CNARN-ARN 177

Db 524 GCCCCAGCCGGAGACAGTACAGGTGGTTCCAGCGGGTTCCAGGGGTTCCAGGGGGATGCCAGCCTGAC 583
            ||||| ||| ::|| | |||| | |: || ||| ||| : | || |
Cp 176 -CCCCANCCNSWNACNARRCANSWRTTNCCNGCNGTNGGRCAYTGNSWNGCDATNSWDAT 118

Db 584 TTTCTTCACCATGGATGACAGCCTGCCTATTGAGCTTCACGAGCATGAGGTCATT 643
            :|||| || :|: : ||::| | |||: |: | |
Cp 117 NSWNCKDATNGTRTCNSWYTCNSWNACNSWYTCRTCNARYTTDATNARCATNARRTCRTT 58

Db 644 AACATGGGTCTGTGTGGAGTAGCCGGGGTGCCGAATGACTTCGAGGCCT 693
            :||||| :||:| || | :|||| | :|
Cp   57 NGCNARNARNGGNCKRTTRTAYTCNGGRTGNCKNACNSWNARNSWNGCYT 8

RESULT   4
ID   PCT-US96-04294-3 STANDARD; DNA; UNC; 1089 BP.
AC   xxxxxx
DT
DE   Sequence 3, Application PC/TUS9604294
CC   Sequence 3, Application PC/TUS9604294
CC   GENERAL INFORMATION:

```
Db    644  AACAUGGGUCUGUGUGGAGUAGCCGGUGUGGCCGGAAUGACUUCGAGGCCU  693
           |||||: |  ||  ||  || |||:|| |: ||||:|  :: ||::||::||::
Cp    57   NGCNARNARNGGNCKRTTRTAYTCNGRTGNCKNACNSWNARNSWNGCYT    8
```

RESULT 5
ID   US-08-738-413B-8 STANDARD; DNA; UNC; 699 BP.
AC
DT
DE   Sequence 8, Application US/08738413B
CC   Sequence 8, Application US/08738413B
CC   Patent No. 5821106
CC   GENERAL INFORMATION:
CC   APPLICANT: CHUNG, Kwang-Hoe
CC   APPLICANT: KOH, You-Seok
CC   APPLICANT: HWANG, Jae-Hoon
CC   APPLICANT: KIM, Doo-Sik
CC   APPLICANT: YUN, Yung-Dae
CC   APPLICANT: MOON, Hong-Mo
CC   TITLE OF INVENTION: A NOVEL cDNA OF DIRECT-ACTING
CC   TITLE OF INVENTION: FIBRINOLYTIC SERINE PROTEASE
CC   NUMBER OF SEQUENCES: 12
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Darby & Darby PC
CC   STREET: 805 Third Avenue
CC   CITY: New York
CC   STATE: New York
CC   COUNTRY: US
CC   ZIP: 10022
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.30
CC   SOFTWARE:
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/738,413B
CC   FILING DATE: October 23, 1996
CC   CLASSIFICATION: 435
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Ludwig, S. Peter
CC   REGISTRATION NUMBER: 25,351
CC   REFERENCE/DOCKET NUMBER: 0136/0C539
CC   TELECOMMUNICATION INFORMATION:
CC   TELEPHONE: 212-527-7700
CC   TELEFAX: 212-753-6237
CC   TELEX: 236687
CC   INFORMATION FOR SEQ ID NO: 8:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 699 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA to mRNA
CC   ORIGINAL SOURCE:
CC   IMMEDIATE SOURCE:
CC   ORGANISM: Agkistrodon halys brevicaudus
CC   CLONE: protease
SQ   SEQUENCE 699 BP; 192 A; 161 C; 167 G; 179 T; 0 OTHER.

Query Match 16.9%; Score 316; DB 3; Length 699;
Best Local Similarity 44.2%; Pred. No. 8.81e-18;
Matches 80; Conservative 26; Mismatches 75; Indels 0; Gaps 0;

```
Db   481  AGAACATTGTGTGCAGGTATCCTGGAGGAGGCAAAGATTCATGTAACGGTGACTCTGGG  540
          ::| | |||:||||  ||    |||:||/:||| |||:: |||:|| ||
Qy   280  WSNATGTTYTGYGCNGGNGNGGNCARNNNGAYWSNTGYAAYGGNGAYWSNSNGGN  339

Db   541  GGACCCCTCATCTGTAATGACAATTCCAGGGCATTGCATTCATATTGGGGCCGATACTTGT  600
          ||/:|| ::|| ||||:|||:|:|||| | | ||| :: ||
Qy   340  GGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGY  399

Db   601  GCCCAACCGCGTGAGCCTGGCCTCACACCAAGGTCTTTGATTATATTGATTGGATCCAA  660
          ||: |    | |:| ||  || ||| ||  |:| | |  |  ||| :: | ||
```

```
Qy   400  GGNCARGTNGGNGTNCCNGNGTNTAYACNAAYYTNGYAARTTYACNGARTGGATHGAR  459
          ||:    ||: |  ||:|||     |||| : || |||:|||||: ||||:|
Db   661  A  661
          |
Qy   460  A  460
```

RESULT 6
ID   US-08-684-862-11 STANDARD; DNA; UNC; 957 BP.
AC
DT
DE   Sequence 11, Application US/08684862
CC   Sequence 11, Application US/08684862
CC   Patent No. 5759541
CC   GENERAL INFORMATION:
CC   APPLICANT: Bach, Alfred
CC   APPLICANT: Hillen, Heinz
CC   APPLICANT: Bialojan, Siegfried
CC   TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
CC   TITLE OF INVENTION: Thereof
CC   NUMBER OF SEQUENCES: 14
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Keil & Weinkauf
CC   STREET: 1101 Connecticut Avenue
CC   CITY: Washington
CC   STATE: D.C.
CC   COUNTRY: USA
CC   ZIP: 20036
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
CC   COMPUTER: IBM AT-compatible, 80286 processor
CC   OPERATING SYSTEM: MS-DOS version 5.0
CC   SOFTWARE: WordPerfect version 5.1
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/684,862
CC   FILING DATE:
CC   CLASSIFICATION: 435
CC   PRIOR APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/361,705
CC   FILING DATE:
CC   PRIOR APPLICATION DATA:
CC   APPLICATION NUMBER: 07/966,040
CC   FILING DATE: 30-DEC-1992
CC   APPLICATION NUMBER: PCT/EP91/01361
CC   FILING DATE: 19-JUL-1991
CC   INFORMATION FOR SEQ ID NO: 11:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 957 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA to mRNA
CC   ORIGINAL SOURCE:
CC   ORGANISM: Agkistrodon rhodostoma
CC   FEATURE:
CC   LOCATION: 210 to 911
CC   OTHER INFORMATION: the coding region shown in (2)(ix)(B)
CC   OTHER INFORMATION: codes for the protein of SEQ ID NO: 6
SQ   SEQUENCE 957 BP; 270 A; 203 C; 222 G; 258 T; 4 OTHER.

Query Match 15.3%; Score 286; DB 3; Length 957;
Best Local Similarity 41.8%; Pred. No. 7.72e-15;
Matches 74; Conservative 27; Mismatches 76; Indels 0; Gaps 0;

```
Db   693  AGAACATTGTGTGCAGGTATCGTCGCAAGGAGGCAAAGATCACATGTATGTGACTCTCGGA  752
          ::| | |||:||||  ||  |   |||:||:||| |||:: ||: ||
Qy   280  WSNATGTTYTGYGCNGGNGNGGNCARNNNGAYWSNTGYAAYGGNGAYWSNSNGGN  339

Db   753  GGACCTCTCATCTGTAATGAACAAGTCCAGGGCCATTGATCTTATGGGGCCGCATCCTTGT  812
          ||| || ::|| ||||:|||::|:|||| | ||| || :: ||
Qy   340  GGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGY  399
```

```
Db   813 GGCCAACCTCTTAAGCCTGGTATCTCACACCAGGCTCCATGATTATAATGACTGGATC 869
         ||.||:      ||||| |.||||| .:| ||||||  .| |  :.| || |||||:
Qy   400 GGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNTGYAARTTYACNGARTGGATH 456
```

RESULT 7
```
ID   US-08-684-862-10 STANDARD; DNA; UNC; 988 BP.
AC   xxxxxx
DT
DE   Sequence 10, Application US/08684862
DE   Sequence 10, Application US/08684862
CC   Patent No. 5759541
CC   GENERAL INFORMATION:
CC      APPLICANT: Bach, Alfred
CC      APPLICANT: Hillen, Heinz
CC      APPLICANT: Bialojan, Siegfried
CC      TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
CC      TITLE OF INVENTION: Thereof
CC      NUMBER OF SEQUENCES: 14
CC      CORRESPONDENCE ADDRESS:
CC         ADDRESSEE: Keil & Weinkauf
CC         STREET: 1101 Connecticut Avenue
CC         CITY: Washington
CC         STATE: D.C.
CC         COUNTRY: USA
CC         ZIP: 20036
CC      COMPUTER READABLE FORM:
CC         MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
CC         COMPUTER: IBM AT-compatible, 80286 processor
CC         OPERATING SYSTEM: MS-DOS version 5.0
CC         SOFTWARE: WordPerfect version 5.1
CC      CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER: US/08/684,862
CC         FILING DATE:
CC         CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER: US/08/361,705
CC         FILING DATE:
CC      PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER: 07/966,040
CC         FILING DATE: 30-DEC-1992
CC         APPLICATION NUMBER: PCT/EP91/01361
CC         FILING DATE: 19-JUL-1991
CC      INFORMATION FOR SEQ ID NO: 10:
CC         SEQUENCE CHARACTERISTICS:
CC            LENGTH: 988 base pairs
CC            TYPE: nucleic acid
CC            STRANDEDNESS: single
CC            TOPOLOGY: linear
CC         MOLECULE TYPE: cDNA to mRNA
CC         ORIGINAL SOURCE:
CC            ORGANISM: Agkistrodon rhodostoma
CC         FEATURE:
CC            LOCATION: 197 to 904
CC            OTHER INFORMATION: the coding region shown in (2)(ix)(B)
CC            OTHER INFORMATION: codes for the protein of SEQ ID NO: 5
SQ   SEQUENCE 988 BP; 290 A; 211 C; 224 G; 258 T; 5 OTHER.
```

```
Query Match          14.2%;  Score 266;  DB 3;  Length 988;
Best Local Similarity 41.2%;  Pred. No. 6.65e-13;
Matches 73;  Conservative 26;  Mismatches 78;  Indels 0;  Gaps 0;

Db   686 AGAATATTGTGTCAGGTGTCCTGCAAGGAGGCATGATACCATGTAATCATGACTCTGGG 745
         | .|::|| | | |||||||::  ||::|||| | :|||||:|:||:  |||::::  ||
Qy   280 WSNATGTTYTGYGCNGGNGGNGGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGN 339

Db   746 GGACCTCTCATCTGTGATGAACAATTCCAGGGCATTGTATCTTGGGGACCCTATCCTTGT 805
         ||  ||| :|:||| | ||| || || ||||| |::| :|| |||:::: |  ||||:||
Qy   340 GGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGY 399

Db   806 GCCCAACCGGCTAACGCTGCCATCTACACCAAAGTCTTCAATTATCTTGTCTGGGTC 862
         | .|||:  |||||:|| |:||| | ||||:  |||||:  :||||| |:|.:||.
Qy   400 GGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNTGYAARTTYACNGARTGGATH 456
```

RESULT 8
```
ID   US-08-358-782D-13 STANDARD; DNA; UNC; 992 BP.
AC   xxxxxx
DT
DE   Sequence 13, Application US/08358782D
DE   Sequence 13, Application US/08358782D
CC   Patent No. 5674682
CC   GENERAL INFORMATION:
CC      APPLICANT: Croce, Carlo
CC      APPLICANT: Gomella, Leonard
CC      APPLICANT: Mulholland, S. Grant
CC      APPLICANT: Moreno, Jose
CC      APPLICANT: Fischer, Rainer
CC      TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
CC      NUMBER OF SEQUENCES: 14
CC      CORRESPONDENCE ADDRESS:
CC         ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
CC         STREET: One Liberty Place 46th. Floor
CC         CITY: Philadelphia
CC         STATE: PA
CC         ZIP: 19103
CC      COMPUTER READABLE FORM:
CC         MEDIUM TYPE: Floppy disk
CC         COMPUTER: IBM PC compatible
CC         OPERATING SYSTEM: PC-DOS/MS-DOS
CC         SOFTWARE: Patentin Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER: US/08/358,782D
CC         FILING DATE: 15-DEC-1994
CC         CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC         NAME: Beardell, Lori Y.
CC         REGISTRATION NUMBER: 34,293
CC         REFERENCE/DOCKET NUMBER: TJU-1327
CC      TELECOMMUNICATION INFORMATION:
CC         TELEPHONE: 215-568-3100
CC         TELEFAX: 215-568-3439
CC      INFORMATION FOR SEQ ID NO: 13:
CC         SEQUENCE CHARACTERISTICS:
CC            LENGTH: 992 base pairs
CC            TYPE: nucleic acid
CC            STRANDEDNESS: single
CC            TOPOLOGY: linear
CC         MOLECULE TYPE: DNA (genomic)
CC         HYPOTHETICAL: NO
CC         ANTI-SENSE: NO
SQ   SEQUENCE 992 BP; 210 A; 291 C; 272 G; 217 T; 2 OTHER.
```

```
Query Match          14.1%;  Score 263;  DB 2;  Length 992;
Best Local Similarity 36.6%;  Pred. No. 1.29e-12;
Matches 98;  Conservative 39;  Mismatches 131;  Indels 0;  Gaps 0;

Db   554 CTTCAGTGTGGACCTCCATGTAATTCCAATGACGTGTGCGCAAGTTCACCCTCAG 613
         :|.||||||:|      | ||| |||||  |||||||||||   ||||||||| 
Qy   208 YTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNTAYGAYCCN 267

Db   614 AAGGTGACCAAGTTCATCGTGCTCGACCGTGGACAGGGGCAAAAGCACCTGCTCG 673
         || | :| | ||:|| || |:||  || :||:|     .   :||
Qy   268 YTNTAYCAYCCWSNATGTTYTGYGCNGGNGGNGGNCARNNNGAYWSNTGYAAY 327

Db   674 GGTGATTCTGGGGGCCCCACTTGTCGTAATGGTGTGCTTCAAGGTATCACGTCATGGGGC 733
         ||  || |:| ||:||| | :|| | || ||||||||  ||:||| | :: |||||
Qy   328 GGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGN 387

Db   734 AGTGAACCATGTGCCCTGCCCGAAAGGCCTTCCCTGTACACCAAGGTGTGCATTACCGG 793
         | | | ||| :  |:    :| ||  |  :||  || | ||  ||:|:| |  |:
Qy   388 AARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNTGYAARTTYACN 447

Db   794 AAGTGGATCAAGGACACCATCGTGGCCA 821
         :||||||  |::  |  | || | |:: 
Qy   448 GARTGGATHGARAARACNGTNCARGCNW 475
```

```
RESULT 9
ID   US-08-358-782D-14 STANDARD; DNA; UNC; 1462 BP.
AC   xxxxxx
DT
DE   Sequence 14, Application US/08358782D
DE   Sequence 14, Application US/08358782D
CC   Patent No. 5674682
CC   GENERAL INFORMATION:
CC      APPLICANT: Croce, Carlo
CC      APPLICANT: Gomella, Leonard
CC      APPLICANT: Mulholland, S. Grant
CC      APPLICANT: Moreno, Jose
CC      APPLICANT: Fischer, Rainer
CC      TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
CC      NUMBER OF SEQUENCES: 14
CC      CORRESPONDENCE ADDRESS:
CC         ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
CC         STREET: One Liberty Place 46th. Floor
CC         CITY: Philadelphia
CC         STATE: PA
CC         ZIP: 19103
CC      COMPUTER READABLE FORM:
CC         MEDIUM TYPE: Floppy disk
CC         COMPUTER: IBM PC compatible
CC         OPERATING SYSTEM: PC-DOS/MS-DOS
CC         SOFTWARE: PatentIn Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER: US/08/358,782D
CC         FILING DATE: 15-DEC-1994
CC         CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC         NAME: Beardell, Lori Y.
CC         REGISTRATION NUMBER: 34,293
CC         REFERENCE/DOCKET NUMBER: TJU-1327
CC      TELECOMMUNICATION INFORMATION:
CC         TELEPHONE: 215-568-3100
CC         TELEFAX: 215-568-3439
CC   INFORMATION FOR SEQ ID NO: 14:
CC      SEQUENCE CHARACTERISTICS:
CC         LENGTH: 1462 base pairs
CC         TYPE: nucleic acid
CC         STRANDEDNESS: single
CC         TOPOLOGY: linear
CC         MOLECULE TYPE: DNA (genomic)
CC         HYPOTHETICAL: NO
CC         ANTI-SENSE: NO
SQ   SEQUENCE 1462 BP; 343 A; 374 C; 422 G; 323 T; 0 OTHER.

Query Match       14.1%;  Score 263;  DB 2;  Length 1462;
Best Local Similarity 36.6%;  Pred. No. 1.29e-12;
Matches  98;  Conservative  39;  Mismatches 131;  Indels   0;  Gaps   0;

Db     535 CTTCAGTGTGTGGACCTCCATGTTATTTCCAATGACGTGTGTGCGCAAGTTCACCCTCAG 594
               | |:||:||  || | | :: || ||  ||   | | : || : :|
Qy     208 YTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNTAYGAYCCN 267

Db     595 AAGGTGACCAAGTTCATGCTGTGTGCTGGACGCTGGACAGGGGCAAAAGCACCTGCTCG 654
               |||: |  ||| :|| |: || |  :: ||
Qy     268 YTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGNGGNCARNNNCARNNNGAYWSNTGYAAY 327

Db     655 GGTGATTCTGGGGGCCCACTTGTCTGTAATGGTGTGCTTCAAGGTATCACGTCATGGGGC 714
               || ||::: || ||| :|| |: ::|:|| : :|||:| | |||:| | :: |
Qy     328 GGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGN 387

Db     715 AGTGAACCATGTGCCCTGCCCGAAAGGCCTTCCCTGTACACCAAGGTGGTCATTACCGG 774
                |  |||| |  | ||   |  ||  |  || :|| || ||  || | |:
Qy     388 AARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNTGYAARTTYACN 447

Db     775 AAGTGGATCAAGGACGACCATCGTGGCCA 802
               |:|||||||:|:|||  |:|:|| : :|
Qy    '448 GARTGGATHGARAAARACNGTNCARGCNW 475
```

```
RESULT 10
ID   US-08-684-862-12 STANDARD; DNA; UNC; 840 BP.
AC   xxxxxx
DT
DE   Sequence 12, Application US/08684862
DE   Sequence 12, Application US/08684862
CC   Patent No. 5759541
CC   GENERAL INFORMATION:
CC      APPLICANT: Bach, Alfred
CC      APPLICANT: Hillen, Heinz
CC      APPLICANT: Bialojan, Siegfried
CC      TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
CC      TITLE OF INVENTION: Thereof
CC      NUMBER OF SEQUENCES: 14
CC      CORRESPONDENCE ADDRESS:
CC         ADDRESSEE: Keil & Weinkauf
CC         STREET: 1101 Connecticut Avenue
CC         CITY: Washington
CC         STATE: D.C.
CC         COUNTRY: USA
CC         ZIP: 20036
CC      COMPUTER READABLE FORM:
CC         MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
CC         COMPUTER: IBM AT-compatible, 80286 processor
CC         OPERATING SYSTEM: MS-DOS version 5.0
CC         SOFTWARE: WordPerfect version 5.1
CC      CURRENT APPLICATION DATA:
CC         APPLICATION NUMBER: US/08/684,862
CC         FILING DATE:
CC         CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER: US/08/361,705
CC         FILING DATE:
CC      PRIOR APPLICATION DATA:
CC         APPLICATION NUMBER: 07/966,040
CC         FILING DATE: 30-DEC-1992
CC         APPLICATION NUMBER: PCT/EP91/01361
CC         FILING DATE: 19-JUL-1991
CC   INFORMATION FOR SEQ ID NO: 12:
CC      SEQUENCE CHARACTERISTICS:
CC         LENGTH: 840 base pairs
CC         TYPE: nucleic acid
CC         STRANDEDNESS: single
CC         TOPOLOGY: linear
SQ   SEQUENCE 840 BP; 238 A; 180 C; 204 G; 217 T; 1 OTHER.

Query Match       13.6%;  Score 255;  DB 3;  Length 840;
Best Local Similarity 41.2%;  Pred. No. 7.52e-12;
Matches  75;  Conservative  25;  Mismatches  82;  Indels   0;  Gaps   0;

Db     606 AGAGTATTGTGTGCAGTGACCTGCGAGGACGCAGAGATTCATGTAATAGTGACTCTGGG 665
               :: |  ||:||||  :||:|| |  | ::: ||::: |||:|| |::::||
Qy     280 WSNATGTTYTGYGCNGGNGGNGGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGN 339

Db     666 GGACCTCTCATTGTAATGAAGAACTCCATGGCATTGTAGCTAGGGGACCCAATCCTTGT 725
               |::: :||:: || | |  :| || | | |   |:: |  ||  |  || |: :||
Qy     340 GGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGY 399

Db     726 GCCCAGCCGAATAAGCCTGCCCTCTACACCAGCGTCTACGATTATCGTGACTGGGTCAAT 785
               |: |  |||| ||| | |  || || |||  | |: || | :|||:| || |:
Qy     400 GGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGAR 459

Db     786 AA 787
               ||
Qy     460 AA 461
```

```
RESULT 11
ID   PCT-US95-06157-7 STANDARD; DNA; UNC; 760 BP.
AC   xxxxxx
DT
```

AC   xxxxxx
DT   Sequence 9, Application PC/TUS9506157
DE   Sequence 9, Application PC/TUS9506157
CC   GENERAL INFORMATION:
CC   APPLICANT: Mayo Foundation for Medical Education
CC   APPLICANT: and Research
CC   APPLICANT: Hybritech Incorporated
CC   APPLICANT: Tindall, Donald J.
CC   APPLICANT: Young, Charles Y.F.
CC   APPLICANT: Saedi, Mohammed S.
CC   TITLE OF INVENTION: Recombinant HK2 Polypeptide
CC   NUMBER OF SEQUENCES: 18
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
CC   STREET: 3500 IDS Center
CC   CITY: Minneapolis
CC   STATE: MN
CC   COUNTRY: USA
CC   ZIP: 55402
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.25
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: PCT/US95/06157
CC   FILING DATE:
CC   CLASSIFICATION:
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Raasch, Kevin W.
CC   REGISTRATION NUMBER: 35,561
CC   REFERENCE/DOCKET NUMBER: 150.148WO1
CC   TELECOMMUNICATION INFORMATION:
CC   TELEPHONE: 612-339-0331
CC   TELEFAX: 612-339-3061
CC   INFORMATION FOR SEQ ID NO: 9:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 766 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: double
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA
CC   FEATURE:
CC   NAME/KEY: CDS
CC   LOCATION: 1..732
SQ   SEQUENCE 766 BP; 162 A; 218 C; 219 G; 167 T; 0 OTHER.

Query Match   13.5%;   Score 253; DB 4; Length 766;
Best Local Similarity 36.6%;   Pred. No. 1.17e-11;
Matches 98; Conservative 38; Mismatches 132; Indels 0; Gaps 0;

Db 460 CTTCAGTGTGTGAGCCTCCATCTCCTGTCCAATGACATGTGCTCCAATGACATGTGCTAGAGCTTACTCTGAG 519
         :|:||::||:|:||| ::| | : | :| ||| : |:| |:::::||:|
Qy 208 YTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNTAYGAYCCN 267

Db 520 AAGGTGACAGAGTTCATGTTGTGTCCTGGCTCTGGACAGGTGGTAAAGACACTTGTGGG 579
         || |:|| :||:|||||| || :||||::|| |||||||||:|| : |||
Qy 268 YTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGNGGNCARNNNGAYWSNTGYAAY 327

Db 580 GGTGATTCTGGGGGTCCACTTGTGCTTCAAGGTATCACATCATGGGGC 639
         || |:|||| : || | || : ||| :| |||: :|
Qy 328 GGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTATGGNTNWSNTTYGGN 387

Db 640 CCTGAGCCATGTGCCCTGCCTGAAAAGCCTGCTGTGTACACCAAGGTGGTGCATTACCGG 699
         ||||| ||:::| |:: :|| :: ||: ||||::: ||:|| |||
Qy 388 AARGCNCCNTGYGCNGGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYTNTGYAARTTYACN 447

Db 700 AAGTGGATCAAGTACACCATCCGCAGCCA 727
         ||:|:||:|| :|||:: :|| | ||:
Qy 448 GARTGGATHGARAARACNGTNCARGCNW 475

RESULT 12
ID   PCT-US95-06157-9 STANDARD; DNA; UNC; 766 BP.

RESULT 13
ID PCT-US95-06157-5 STANDARD; DNA; UNC; 832 BP.
AC xxxxxx
DT
DE Sequence 5, Application PC/TUS9506157
DE Sequence 5, Application PC/TUS9506157
CC GENERAL INFORMATION:
CC   APPLICANT: Mayo Foundation for Medical Education
CC   APPLICANT: and Research
CC   APPLICANT: Hybritech Incorporated
CC   APPLICANT: Tindall, Donald J.
CC   APPLICANT: Young, Charles Y.F.
CC   APPLICANT: Saedi, Mohammed S.
CC   TITLE OF INVENTION: Recombinant HK2 Polypeptide
CC   NUMBER OF SEQUENCES: 18
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
CC   STREET: 3500 IDS Center
CC   CITY: Minneapolis
CC   STATE: MN
CC   COUNTRY: USA
CC   ZIP: 55402
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Floppy disk
CC   COMPUTER: IBM PC compatible
CC   OPERATING SYSTEM: PC-DOS/MS-DOS
CC   SOFTWARE: PatentIn Release #1.0, Version #1.25
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: PCT/US95/06157
CC   FILING DATE:
CC   CLASSIFICATION:
CC   ATTORNEY/AGENT INFORMATION:
CC   NAME: Raasch, Kevin W.
CC   REGISTRATION NUMBER: 35,561
CC   REFERENCE/DOCKET NUMBER: 150.148WO1
CC   TELECOMMUNICATION INFORMATION:
CC   TELEPHONE: 612-339-0331
CC   TELEFAX: 612-339-3061
CC INFORMATION FOR SEQ ID NO: 5:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 832 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: double
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA
CC   FEATURE:
CC   NAME/KEY: CDS
CC   LOCATION: 10..792
SQ SEQUENCE 832 BP; 169 A; 237 C; 241 G; 185 T; 0 OTHER.

Query Match   13.5%; Score 253; DB 4; Length 832;
Best Local Similarity 36.6%; Pred. No. 1.17e-11;
Matches 98; Conservative 38; Mismatches 132; Indels 0; Gaps 0;

Db 520 CTTCAGTGTGTGAGCCCCATCTCCTGTCCAATGACATGTGTGCTAGAGCTTACTCTGAG 579
         :|:||||||:||:|: | || |||:|| || | |||| :| : ||: :| |
Qy 208 YTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNTAYGAYCCN 267
Db 580 AAGGTGACAGAGTTCATGTTGTGTGGGCTGGGCTTCTGGACAGGTGGTAAAGACACTGTGGG 639
         || || | |||||||:|||||||||||||||| | | |||||||| || | | ||
Qy 268 YTNTAYCAYCCNWSNAATGTTYTGYGCNGGNGGNGGNCARNNNCARNNNGAYWSNTGYAAY 327
Db 640 GGTGAGATTCTGGGGGTCCACTTGTCTGTAATGTGTGTCTTCAAGGTATCACATCATGGGGC 699
         ||:|| || | ||||| |:|| |:| ||||| | |||:|||| | || ||| | |||
Qy 328 GGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGN 387
Db 700 CCTGAGCCATGTGCCCCTGCCTGAAAAGCCTCGTGTGTACACCAAGGTGGTGCATTACCGG 759
         || | |||||| || |||||||| ||||||||| || ||:||:|||:|||| ||
Qy 388 AARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNGTAARTTYACN 447
Db 760 AAGTGGATCAAGTAACACCATCCGCAGCCA 787
         ||||||||:||:|| || || || |
Qy 448 GARTGGATHGARAAARACNGTNCARGCNW 475

RESULT 14
ID US-08-684-862-8 STANDARD; DNA; UNC; 1096 BP.
AC xxxxxx
DT
DE Sequence 8, Application US/08684862
DE Sequence 8, Application US/08684862
DE Patent No. 5759541
CC GENERAL INFORMATION:
CC   APPLICANT: Bach, Alfred
CC   APPLICANT: Hillen, Heinz
CC   APPLICANT: Bialojan, Siegfried
CC   TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
CC   TITLE OF INVENTION: Thereof
CC   NUMBER OF SEQUENCES: 14
CC   CORRESPONDENCE ADDRESS:
CC   ADDRESSEE: Keil & Weinkauf
CC   STREET: 1101 Connecticut Avenue
CC   CITY: Washington
CC   STATE: D.C.
CC   COUNTRY: USA
CC   ZIP: 20036
CC   COMPUTER READABLE FORM:
CC   MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
CC   COMPUTER: IBM AT-compatible, 80286 processor
CC   OPERATING SYSTEM: MS-DOS version 5.0
CC   SOFTWARE: WordPerfect version 5.1
CC   CURRENT APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/684,862
CC   FILING DATE:
CC   CLASSIFICATION: 435
CC   PRIOR APPLICATION DATA:
CC   APPLICATION NUMBER: US/08/361,705
CC   FILING DATE:
CC   PRIOR APPLICATION DATA:
CC   APPLICATION NUMBER: 07/966,040
CC   FILING DATE: 30-DEC-1992
CC   APPLICATION NUMBER: PCT/EP91/01361
CC   FILING DATE: 19-JUL-1991
CC INFORMATION FOR SEQ ID NO: 8:
CC   SEQUENCE CHARACTERISTICS:
CC   LENGTH: 1096 base pairs
CC   TYPE: nucleic acid
CC   STRANDEDNESS: single
CC   TOPOLOGY: linear
CC   MOLECULE TYPE: cDNA to mRNA
CC   ORIGINAL SOURCE: Agkistrodon rhodostoma
CC   FEATURE:
CC   LOCATION: 144 to 841
CC   OTHER INFORMATION: the coding region shown in (2)(ix)(B)
CC   OTHER INFORMATION: codes for the protein of SEQ ID NO: 3
SQ SEQUENCE 1096 BP; 310 A; 227 C; 247 G; 312 T; 0 OTHER.

Query Match   13.4%; Score 250; DB 3; Length 1096;
Best Local Similarity 42.0%; Pred. No. 2.25e-11;
Matches 73; Conservative 23; Mismatches 78; Indels 0; Gaps 0;

Db 636 ATATTGTGCAGGTGACCTGCCAAGGACGCCTAGAATTCATGTCACTGTGACTGCTGGGGGA 695
         | ||||||||:||||| |||||||| || |||||||||||:|| |||||| ||:| ||
Qy 283 ATGTTYTGYGCNGGNGGNGGNGGNCARNNNCARNNNGAYWSNGYAAYGNGAYWSNGGNGGN 342
Db 696 CCTCTCATTGTAGTGAAGAATTCCATGGCATTGTATATCGGGGACCCCAATCCTTGTGCC 755
         ||:||:||||| ||||| |||| ||| |||| ||||| |||||| |||||||| |||||||
Qy 343 CCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGN 387
Db 756 CAACCAGATAAGCCTGCCCTCACACCAACATCTTCGATCATCTTCACTGGATC 809
         || || || ||||||| || || || ||||| |||| || |||||||||||
Qy 403 CARGTNGGNGTNCCNGGNGTNTAYACNAAYYTNGTAARTTYACNGARTGGATH 456

RESULT 15

Search completed: Wed Sep 29 02:17:32 1999
Job time : 36 secs.

```
ID   US-08-684-862-9 STANDARD; DNA; UNC; 1333 BP.
AC   xxxxxx
DT
DE   Sequence 9, Application US/08684862
DE   Sequence 9, Application US/08684862
CC   Patent No. 5759541
CC   GENERAL INFORMATION:
CC     APPLICANT: Bach, Alfred
CC     APPLICANT: Hillen, Heinz
CC     APPLICANT: Bialojan, Siegfried
CC     TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
CC     TITLE OF INVENTION: Thereof
CC     NUMBER OF SEQUENCES: 14
CC     CORRESPONDENCE ADDRESS:
CC       ADDRESSEE: Keil & Weinkauf
CC       STREET: 1101 Connecticut Avenue
CC       CITY: Washington
CC       STATE: D.C.
CC       COUNTRY: USA
CC       ZIP: 20036
CC     COMPUTER READABLE FORM:
CC       MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
CC       COMPUTER: IBM AT-compatible, 80286 processor
CC       OPERATING SYSTEM: MS-DOS version 5.0
CC       SOFTWARE: WordPerfect version 5.1
CC     CURRENT APPLICATION DATA:
CC       APPLICATION NUMBER: US/08/684,862
CC       FILING DATE:
CC       CLASSIFICATION: 435
CC     PRIOR APPLICATION DATA:
CC       APPLICATION NUMBER: US/08/361,705
CC       FILING DATE:
CC     PRIOR APPLICATION DATA:
CC       APPLICATION NUMBER: 07/966,040
CC       FILING DATE: 30-DEC-1992
CC       APPLICATION NUMBER: PCT/EP91/01361
CC       FILING DATE: 19-JUL-1991
CC   INFORMATION FOR SEQ ID NO: 9:
CC     SEQUENCE CHARACTERISTICS:
CC       LENGTH: 1333 base pairs
CC       TYPE: nucleic acid
CC       STRANDEDNESS: single
CC       TOPOLOGY: linear
CC     MOLECULE TYPE: cDNA to mRNA
CC     ORIGINAL SOURCE:
CC       ORGANISM: Agkistrodon rhodostoma
CC     FEATURE:
CC       LOCATION: 231 to 935
CC       OTHER INFORMATION: the coding region shown in (2)(1x)(B)
CC       OTHER INFORMATION: codes for the protein of SEQ ID NO: 4
SQ   SEQUENCE 1333 BP; 427 A; 306 C; 306 G; 316 T; 17 OTHER.
```

```
Query Match          12.0%;  Score 225;  DB 3;  Length 1333;
Best Local Similarity 39.2%;  Pred. No. 5.06e-09;
Matches  71; Conservative 25; Mismatches 85; Indels 0; Gaps 0;

Db  717 AGAACATTATGTGCTGGTATCCTGGAAGGAGGCAAAAGTGCATGTGACGGTGACTCTGG 776
             :: | || || ||| || || |         |: || 1:|| ||::: ||
Qy  280 WSNATGTTYTGYGCNGNGGGNGGNGGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGN 339

Db  777 GGACCCCTCAACTGTAATGGAGAAATCCAGGGCATTGTATCTTGGGGGGTAATATTTGT 836
            |||| :) : |||:|||:|| | || || || || ||           |):
Qy  340 GGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGY 399

Db  837 GCTCAACCGGCTAAGCCTGCCCACTACNCCAGGTGCCGATTATACTGATTGGATTAAG 896
            - | || - ||| - ||  | ||  ||| ||  || ||::||| |||||::| ::
Qy  400 GGNCARGTNGGNGTNCCNGNGTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGAR 459

Db  897 A 897
        -
Qy  460 A 460
```

This Page Blank (uspto)

OM of: US-09-030-606-173 to: Issued_Patents_AA:*   out_format : pfs

Date: Sep 25, 1999   4:01 PM

About: Results were produced by the GenCore software, version 4.5,
      Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frame_n2p model -DEV=xlp
 -Q=/cgn2.1/USPTO.spool/US09030606/runat_24091999_171616_29813/app_query.fasta.1
 -DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=1.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
 -NCPU=3 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-173
Query length: 1265
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1 +	1081.50	1539.19	4.8e-79	248	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2 +	519.50	734.70	3.1e-34	253	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2 +	519.50	734.70	3.1e-34	253	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-12 +	505.00	715.13	4.2e-33	225	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-11 +	449.50	635.18	1.1e-28	237	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11 +	449.50	635.18	1.1e-28	237	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:PCT-US95-06157-1 +	448.50	633.70	1.4e-28	238	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8 +	448.50	633.70	1.4e-28	238	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-10 +	448.50	633.45	1.4e-28	244	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-5 +	448.50	632.78	1.4e-28	261	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-6 +	448.00	632.78	1.5e-28	244	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-361-395-1 +	447.50	632.74	1.5e-28	237	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-10 +	447.50	632.32	1.6e-28	237	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-10 +	447.50	632.32	1.6e-28	237	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:PCT-US95-06157-16 +	447.50	632.32	1.6e-28	240	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-228A-1 +	444.00	627.18	3.1e-28	240	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-09303-1 +	444.00	627.18	3.1e-28	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-7 +	443.00	626.27	3.7e-28	223	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-7 +	443.00	626.27	3.7e-28	228	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-7 +	443.00	626.27	3.7e-28	228	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-7 +	443.00	626.27	3.7e-28	228	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-844-024-2 +	442.50	625.16	4.1e-28	237	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-9 +	438.00	619.33	9.3e-28	223	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-9 +	438.00	619.33	9.3e-28	223	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9 +	438.00	619.33	9.3e-28	223	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-9 +	438.00	619.33	9.3e-28	223	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-3 +	437.50	617.15	1.1e-27	258	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-8 +	426.00	600.54	8.8e-27	232	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-8 +	426.00	600.54	8.8e-27	232	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-790-137-1 +	426.00	600.54	8.8e-27	262	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-3 +	426.00	600.54	8.8e-27	262	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-681-151-4 +	426.00	600.54	8.8e-27	262	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-4 +	416.50	586.90	5.0e-26	263	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-9 +	416.00	585.34	5.6e-26	286	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-9 +	416.00	585.34	5.6e-26	286	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-9 +	416.00	585.34	5.6e-26	286	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-8 +	397.00	560.26	1.7e-24	232	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-8 +	397.00	560.26	1.7e-24	232	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-8 +	397.00	560.26	1.7e-24	232	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-8 +	397.00	560.26	1.7e-24	232	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-7 +	394.00	554.04	3.2e-24	281	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-7 +	394.00	554.04	3.2e-24	281	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-7 +	394.00	554.04	3.2e-24	281	

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1

/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-10 +   393.50   553.69   3.5e-24   271
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-10 +   393.50   553.69   3.5e-24   271
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-10 +   393.50   553.69   3.5e-24

seq_documentation_block:
; Sequence 1; Application US/0874026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-08-744-026-1

alignment_scores:
     Quality: 1081.50   Length: 243
       Ratio: 5.077   Gaps: 2
Percent Similarity: 87.654   Percent Identity: 86.008

alignment_block:
US-09-030-606-173 x US-08-744-026-1

Align seg 1/1   to: US-08-744-026-1   from: 1   to: 248

   10 CACTCGGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG   59
        |||||||||||||||||||||||||||||||||||||||||||||||||||
   34 HisSerGlnProTrpGlnAlaAlaLeuValMetGluAsnGluLeuPheCy   50

   60 CTCGGGCGTCCTGGTCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTT   109
        |||||||||||||||||||||||||||||||||||||||||||||||||
   50 sSerGlyValLeuValHisProGlnTrpValLeuSerAlaAlaHisCysP   67

  110 TCCAGAACTCCTACACCATCGGCGTGGGCCTGCCAGTCTTCAGGCCCGAC   159
        ||||||||||||||||||||||||||||||||||||||||||||||||||

CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/557,146
;   FILING DATE: 14-DEC-1995
;   CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
;   NAME: Sterner, Richard J.
;   REGISTRATION NUMBER: 35,372
;   REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 819-8783
;   TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 253 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-557-146-2

alignment_scores:
        Quality: 519.50   Length: 247
        Ratio: 3.038   Gaps: 5
Percent Similarity: 69.231   Percent Identity: 42.915

alignment_block:
US-09-030-606-173 x US-08-557-146-2

Align seg 1/1 to: 1 to: 253 US-08-557-146-2 from: 1 to: 253

```
  1 GGCAGCCCG........CACTGCAGCCCTGGCAGGGGCACTGGTCAT  41
    ||:::|||          ||:::|||||||:::||||||
 33 GlyAlaProCysAlaArgGlySerHisProTrpGlnValAlaLeuLeuSe  49

 42 GGGAAACGAATTGTTCGCTCGGGTCCTGGTCGTCATCCGCAGTGGGTGC  91
    ||:::|||  |||:::||  ||||:::|||||||:::.   ::||||||||
 49 rGlyAsnGlnLeuHisCysGlyGlyValLeuValAsnGluArgTrpValL  66

 92 TGTCAGCCGCCACACTGTTCCAGAACTCCTACACCATCGGGCTGGCCCTG 141
    ||:::||||||||||   ::|||:::||||||:::.  |||||
 66 euThrAlaAlaHisCysLysMetAsnGluTyrThrValHisLeuGlySer  82

142 CACAGTCTTGAGGCGCACCAAGAGCCAGGAGCCAGATGGTGAGGCCAG   191
    :::|||  :||||||||  |||:::||||||
 83 AspThrLeu..GlyAspArgArg.......AlaGlnArgIleLysAlaSe  96

192 CCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTCGCTAACGACC 241
    ||||||||||:::||||||||  |||:::   :::|||||||||
 96 rLysSerPheArgHisProGlyTyrSerThrGlnThrHisValAsnAspL 113

242 TCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGG 291
    |||||||||::||||||:::||||:::||   :::.  :::::|
113 euMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSerMetValLys 129

292 AGCATCAGCAATTGCTTCGCAGTGCCCTACCGCGGGGAACTCTTGCCTCGT 341
    :::.   :::   ||||||:::||||||:::||||:::||
130 LysValArgLeuProSerArgCysGluProProGlyThrThrCysThrVa 146

342 TTCTGGCTGGGGTCTGCTGCGGCGAACGGTGAGCTCACGGGTGTGTCTGC 391
    ||||||||||
146 lSerGlyTrpGly.................................... 150

392 CCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGACCCAGAGCTCT 441
    |||   :::|||  |||:::||||||:::|||
151 ................................ThrThrThrSer       154

442 GGGTCCCAGGAGAATGCCTACCGTGCTGCTGAGTGGTCAGTGCGTCGGGTG 491
    :::.   :::   :::|||||||:::||:::||
155 ProAspValThr..PheProSerAspLeuMetCysValAspValLysLeu 170

492 GTGGTCTGAGGAGGTCTCAGTAAGCTCTATGACCCGCTGTACCACCCCAG 541
    :::|||   :::   |||:::||||||   ::: ||
171 IleSerProGlnAspCysThrLysValTyrLysAspLeuLeuGluAsnSe 187
```

```
 67 heGlnAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGluAlaAsp  83

160 CAAGAGCCAGGGAGCCAGAATGGTGAGGCCAGCCTCCGTACGGCACCC   209
    |||||||||||||||||||||||||||||||||||||||||||
 84 GlnGluProGlySerGlnMetValGluAlaSerLeuSerValArgHisPr 100

210 AGAGTACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGG 259
    |||||||||||||||||||||||||||||| ||||||||
100 oGluTyrAsnArgProLeuLeuAlaAsnAspLeuMet***IleLysLeuA 117

260 ACGAAATCCGTGTCCGAGTCTGACACCATCAGCAGTTGCTTCG        309
    |||||||||||||||:::|||||||:::|||
117 spGluSerValSerGlySerAspAsnIleArg***IleSerIle***Ser 133

310 CAGTGCCCTACCGCGGGGAACTCTTGCCTTCTGGCTGGGGTCTGCT     359
    |||||||||||||||||||||||||||||||
134 GlnCysProThrAlaGlyAsnPheCysLeuValSerGlyTrpGlyLeuLe 150

360 GGGCGAACGGTGAGCTCACGGGTGTGTCTGCCCCTCTTCAAGGAGGTCCT 409
    |
150 u.................................... 150

410 CTGCCCAGTCGCGGGGGCTGACGAGCTGCTCTGCCCCAGGCCAGCAATGC 459
    ||:::   |||||
151 ..............AlaAsn..GlyArgMetPro 156

460 CTACCGTGCTGCTGCAGTGCGTGAACGTGTCCGGTGGTGTCTGAGGAGGTCTGC 509
    ||||||||||||||||||||||||||||||||
156 roThrValLeuGlnCysValAsnValSerValValSerGluGluValCys 172

510 AGTAAGCTCTATGACCCGCTGTACCACCCCAGCCATGTTCTGCCGCCGGCGG 559
    ||||||||||||||||||||||||||||||||||
173 SerLysLeuTyrAspProLeuTyrHisProSerMetPheCysAlaGlyGl 189

560 AGGGCAAGACCAGAAGGACTCCTCCAACGGTGACGTGGTGTCTGGGGGGCCCTGA 609
    ||||||||||||||||||||||||||||||||||||||
189 yGlyLysGlnLysAspSerCysAsnGlyAspSerGlyGlyProLeuI   206

610 TCTGCAACGGTACTTGCGAGGGCCTTGTGCTTTCGGAAAAGCCCGTGT   659
    ||||||||||||||||||||||||||||||||||||
206 leCysAsnGlyThrLeuLeuGlyLeuValSerPheGlyLysAlaProCys 222

660 GGGCCAAGTTGGGGTGCCAGGTGTCTACACCAACCTCTGCAAAATTCACTGA 709
    |||||||||||||||||||||||||||||||||||
223 GlyGlnValGlyValProGlyValTyrThrAsnLeuCysLysPheThrGl 239

710 GTGGATAGAGAAAACCGTTCCAGGCCAGT 737
    ||||||||||||||||||
239 uTrpIleGluLysThrValGlnAlaSer 248
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2

```
  1 GGCAGCCCG........CACTCGCAGCCCTGGCAGGCGGCGGCACTGGTCAT 41
                     |||::|||::||||:||||:||:::|||::
 33 GlyAlaProCysAlaArgGlySerHisProTrpGlnValAlaLeuLeuSe 49

 42 GAAAACGAATTGTTCTGCTGGGCGCCTGTGCATCCGCAGTGGGTGC 91
    |||||::||:||:||||| |||::|:||:||||||||| ::|:||||
 49 rGlyAsnGlnLeuHisCysGlyGlyValLeuValAsnGluArgTrpVal 66

 92 TGTCAGCCGCACACTGTTCCAGAACTCTACCACCATCGGCGTCGGGCCTG 141
    ||::||||||||| :::|||::|||| ||::::|||::  ||||||
 66 euThrAlaHisCysLysMetAsnGluTyrThrValHisLeuGlySer 82

142 CACAGTCTTGAGGCGACCAACAGAGCCAGGAGCCAGATGGTGGAGGCCAG 191
    ::|||   :::||||::||||   |||||  :::||||||| ::||:||||
 83 AspThrLeu...GlyAspArgArg......AlaGlnArgIleLysAlaSe 96

192 CCTCTCCGTACGGCACCCAGATACAACAGACACCTTGCTCCGTCAACGACC 241
    |  |||  ||||  ||||||||| ::|||:||:||
 96 rLysSerPheArgHisProGlyTyrSerThrGlnThrHisValAsnAspL 113

242 TCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCGACCACCATCCGG 291
    ||||||::||||||||| |||::::::::::  ||||:  ::|:::
113 euMetLeuValLysLeuAsnSerGlnAlaArgLeuSerMetValLys 129

292 AGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAACTCTTGCCTCGT 341
    ::::|| :::|||::||   |||||||:|||||   |||:|||
130 LysValArgLeuProSerArgCysGluProProGlyThrThrCysThrVa 146

342 TTCTGGCTGGGTCTCTGCTGGGCAACGGTGAGCTCAGGGTGTGTGTCTGC 391
    |||||||| :|||||||     |||::||||||||||||:::|||||
146 lSerGlyTrpGly.................................... 150

392 CCTCTTCAAGGAGGTCCTCTGCCAGTCGCGGGGGGCCGACCCAGAGACTCT 441
                                          ||| |||||
151 ....................................ThrThrThrSer 154

442 GGGTCCAGGCAGAATCCTACCGTGCTGCAGTGCGTGAACGTGTCCGTG 491
    ::: :||||:|||||| ||::|||:  |||||::|||:::||::
155 ProAspValThr..PheProSerAspLeuMetCysValAspValLysLeu 170

492 GTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAG 541
    :::::|| ::|||::|||   :::| :::|||  |||  ::|
171 IleSerProGlnAspCysThrLysValTyrLysAspLeuLeuGluAsnSe 187

542 CATGTTCTGCGCCGGCGAGGCAAGACCAAGAAGGACTCCTGCAACGGTG 591
    |||::||||||:||  ||::::|||| |||::|| ||:|||||
187 rMetLeuCysAlaGlyIleProAspSerLysAsnAlaCysAsnGlyA 204

592 ACTCTGGGGGGCCCCTGATCTCCAACGGTACTTGCAGGGCCTTGTGTCT 641
    |||||||||||||::::|| :|||||||||||:|||||||||||||
204 spSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuValSer 220

642 TTCGGAAAAGCCCCGTGTGGCCAAGTTGGCGTGCTCCACCAGTGTCTACCACCAA 691
    :::|||||||||||:::| |||   ||  :::|||||||||::::|||
221 TrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThrGl 237

692 CCTCTGCAAATTCACTGAGTGGATAGAGAAAACCGTCCAG 731
    :::||||||||||:|::|||:::|||||||
237 nValCysLysPheThrLysTrpIleAsnAspThrMetLys 250
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-12

```
542 CATGTTCTGCGCCGGCGGAGGGCAAGACCAGAAGGACTCCTGCAACGGTG 591
    |||::|||||||
187 rMetLeuCysAlaGlyIleProAspSerLysAsnAlaCysAsnGlyA 204

592 ACTCTGGGGGGCCCCTGATCTCCAACGGTACTTGCAGGGCCTTGTGTCT 641
    |||||||||||
204 spSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuValSer 220

642 TTCGGAAAAGCCCCGTGTGGCCAAGTTGGCGTGCTCCACCAA 691
    ::::|||||||||||
221 TrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThrGl 237

692 CCTCTGCAAATTCACTGAGTGGATAGAGAAAACCGTCCAG 731
    :::||||
237 nValCysLysPheThrLysTrpIleAsnAspThrMetLys 250
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2

alignment_scores:
        Quality: 519.50   Length: 247
          Ratio: 3.038   Gaps: 5
Percent Similarity: 69.231   Percent Identity: 42.915

alignment_block:
US-09-030-606-173 x PCT-US96-04294-2   ..

Align seg 1/1 to: PCT-US96-04294-2 from: 1 to: 253

```
  ADDRESSEE:  White & Case, Patent Department
  STREET:  1155 Avenue of the Americas
  CITY:  New York
  STATE:  New York
  COUNTRY:  U.S.A.
  ZIP:  10036-2787
COMPUTER READABLE FORM:
  MEDIUM TYPE:  Floppy disk
  COMPUTER:  IBM PC compatible
  OPERATING SYSTEM:  PC-DOS/MS-DOS
  SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
  APPLICATION NUMBER:  US/08/557,146
  FILING DATE:  14-DEC-1995
  CLASSIFICATION:  424
ATTORNEY/AGENT INFORMATION:
  NAME:  Sterner, Richard J.
  REGISTRATION NUMBER:  35,372
  REFERENCE/DOCKET NUMBER:  1103326-181
TELECOMMUNICATION INFORMATION:
  TELEPHONE:  (212) 819-8783
  TELEFAX:  (212) 354-8113
SEQUENCE CHARACTERISTICS:
  LENGTH:  225 amino acids
  TYPE:  amino acids
  STRANDEDNESS:  single
  TOPOLOGY:  linear
  MOLECULE TYPE:  polypeptide
US-08-557-146-12
```

```
alignment_scores:
        Quality:  505.00          Length:  248
          Ratio:  2.953            Gaps:  6
Percent Similarity:  68.952   Percent Identity:  42.339

alignment_block:
US-09-030-606-173 x US-08-557-146-12

Align seq 1/1  to: US-08-557-146-12  from: 1  to: 225

  1 GGCAGCCCG........CACTCGCAGCCCTGGCAGGCGGCACTGGTCAT  41
    ||:::|||           ||:::||||||||||||:::
  4 GlyAlaProCysAlaArgGlySerHisProTrpGlnValAlaLeuLeuSe  20

 42 GGAAAACGAATTG...TTCTGCTCGGGCGTCCTGGTGCATCCGCAGTGGG  88
    |||:||| ||||||||       |||||||||::: ::|||
 20 rGlyAsnGlnLeuHisHisCysGlyValLeuValAsnGluArgTrpV  37

 89 TGCTGTCAGCCGCACACTGTTTCCAGAACTCCTACACCATCGGGCTGGGC  138
    ||||:||||||||||||   ::|||:||||||:::|||||:: |||||
 37 alLeuThrAlaAlaHisCysGluThrSerMetAsnGluTyrThrValHisLeuGly  53

139 CTGCACAGTCTTGAGCGCGACCAAGAGCCAGGAGCCAGATGGTGAGGC  188
    ::|||    ||:::||||      :::||    ::::||
 54 SerAspThrLeu..GlyAspArgArg.....AlaGlnArgIleLysAl  67

189 CAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTCGCTAACG  238
    |||| ||| ||||||||       ::|||
 67 aSerLysSerPheArgHisProGlyTyrSerThrGlnThrHisValAsnA  84

239 ACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATC  288
    ||||||||||||:|||:|||::|||  |||: ::
 84 spLeuMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSerMetVal  100

289 CGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACTCTTGCCT  338
    |||||||||:|||||||||||||||:::|||
101 LysLysValArgLeuProSerArgCysGluProProGlyThrThrCysTh  117

339 CGTTTCTGCTGGGTCGTCTGGCGAACGGTGAGCTCACGGGTGTGTGTC  388
    ||||||||||||||
117 rValSerGlyTrpGly.................................  122
```

```
389 TGCCCTCTTCAAGGAGGTCCTTCGCTGCCCAGTCGCGGGGGGCTGACCCAGAGC  438
    |||              :::
123 ....................................ThrThrThr  125

439 TCTGCGTCCCAGGCAGAATGCCTACCGTGCTCAGTGCGTGAACGTGTCG  488
    |||  ::: ::: :||||::  ::|||:::||||::||:::||::::
126 SerProAspValThr..PheProSerAspLeuMetCysValAspValLys  141

489 GTGGTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCCGTGTACCACCC  538
    ::::|||   ||:::||||||:::|||       :::
142 LeuIleSerProGlnAspCysThrGluValTyrLysAspLeuGluGluAs  158

539 CAGCATGTTCTGCGCCGGCAGGGCAAGACCAGAAGGACTCCTGCAACG  588
    ||||:::||||||||   :::::::|||||||::::|||||
158 nSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAsnG  175

589 GTGACTCTGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTG  638
    |||||||||||||||| ::|||:::|| ||||||||||||
175 lyAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuVal  191

639 TCTTTCGGAAAGCCCCGTGCTGGCCAAGTTGGCTGCCAGGTGTCTACAC  688
    ||||:||||||||||||| ::: ||||||||||
192 SerTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrTh  208

689 CAACCCTCGCAAATTCACTGAGTGGGATAGAGAAAACCGTCCAG  731
    |:::::||||||||||||::::||||::: |||:::::
208 rGlnValCysLysPheThrLysTrpIleAsnAspThrMetLys  222
```

Wed Sep 29 14:27:13 1999

US-08-096-946-11

alignment_scores:
        Quality: 449.50   Length: 254
        Ratio: 2.809   Gaps: 5
Percent Similarity: 62.992   Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x US-08-096-946-11

Align seg 1/1 to: US-08-096-946-11 from: 1 to: 237

     10 CACTCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAACGAATTGTTCTG 59
        ||||||||||||||||||::   ::::: :::::::: |
     10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgGlyAlaValCy 26

     60 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 109
        |:::||||||||||||||||||||::::::|||||||||||||||
     26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43

    110 TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC 159
        :::::||:::: :::||| ||||||| |||||||||
     43 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59

    160 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCC 209
        :::::: ||| |||::::||| || ::::|||| | ||||
     60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75

    210 AGAGTACAACAGACCCTTGCTCGCTAAC................ 237
        | :::::: |||
     75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92

    238 ........GACCTCATGCTCATCAAGTGGACGAATCCGTGTCCGAG 276
        ||||||||::::||:::::||| :::::
     92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108

    277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGG 326
        ::||::|||| ::::::: :::::||| |||:::: |
    109 ThrAspAlaValLysValMetAspLeuProThrGlnLeuProAlaLeuGl 125

    327 GAACTCTTGCCTCGTTTCTGGCTGGGGTTGCTGCCGCGAACGGTGAGCTCA 376
        :::::||| :|||||||||||| :::::|||
    125 yThrThrCysTyrTyrAlaSerGlyTrpGlySerIle..Glu......... 137

    377 CGGGTGTGTCTGCCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGG 426
        ::::::|||:::|||||||||||||::||||::|::::::
    137 137

    427 CTGACCCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGC 476
        |||||| :: || ||||||||
    138 ....ProGluGluPheLeu..........ThrProLysLysLeuGlnCys 149

    477 GTGAACGTGTCGGTGGTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCC 526
        ||||||||||||||::::||||||:::||||||||||||
    150 ValGlnLeuHisValIleSerAsnAspValCysAlaGlnValHisProGl 166

    527 GCTGTACCCACCCCAGCATGTTCTGCCGCGCGAGGGCAAGACCAGAAGG 576
        :||:::||||||||||:: |||:||::||||||||||
    166 nLysValThrLysPheMetLeuCysAlaGlyArgTrpThrGlyGlyLyss 183

    577 ACTCCTGCAACGGTGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTG 626
        ::::|::||||||||||||||||||||||:::|||||||
    183 erThrCysSerGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199

    627 CAGGGCCTTGCTTGTGGTTCGGAAAAGCCCGTGTGGCCAAGTTGGCGTGCC 676
        ||||||||::::::|| ||||||||||:::||||||
    200 GlnGlyLeuThrSerTrpGlySerGluProCysAlaLeuProGluArgPr 216

    677 AGGTGTCTACCACCACCTGCAAATTCACTGAGTGGATAGAGAAAACCG 726
        :::::|||:: ||::::::|||| || :::::||| ::::: |||
    216 oSerLeuTyrThrLysValHisTyrArgLysTrpIleLysTyrAspThrI 233

---

     727 TCCAGCCAGT 737
         :: |||::
     233 leValAlaAsn 236

alignment_scores:
        Quality: 449.50   Length: 254
        Ratio: 2.809   Gaps: 5
Percent Similarity: 62.992   Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x PCT-US94-07329-11

Align seg 1/1 to: PCT-US94-07329-11 from: 1 to: 237

     10 CACTCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAACGAATTGTTCTG 59
        |||||||||||||||||||::   ::::: :::::::: |
     10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgGlyAlaValCy 26

     60 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 109
        |:::||||||||||||||||||||::::::|||||||||||||||
     26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43

    110 TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC 159
        :::::||:::: :::||| ||||||| |||||||||
     43 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59

    160 CAAGAGCCAGGAGCCAGATGGTGAGGCCAGCCTCTCCGTACGGCACCC 209

```
COMPUTER:        IBM PC compatible
OPERATING SYSTEM:  PC-DOS/MS-DOS
SOFTWARE:        PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER:    PCT/US95/06157
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME:  Raasch, Kevin W.
REGISTRATION NUMBER:  35,561
REFERENCE/DOCKET NUMBER:  150.148WO1
TELECOMMUNICATION INFORMATION:
TELEPHONE:  612-339-0331
TELEFAX:    612-339-3061
INFORMATION FOR SEQ ID NO:   1:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE:  amino acid
STRANDEDNESS:  single
TOPOLOGY:  linear
MOLECULE TYPE:  peptide
PCT-US95-06157-1
```

```
alignment_scores:
Quality:     449.50           Length:      254
Ratio:       2.809            Gaps:          5
Percent Similarity: 62.992   Percent Identity: 37.402
```

```
alignment_block:
us-09-030-606-173 x PCT-US95-06157-1
```

```
Align seg 1/1  to: PCT-US95-06157-1  from: 1  to: 237

10 CACTCGCAGCCCTGCAGGCGGCACTGGTCATGGAAAACGAATTGTCTG 59
   ||||||||||||||||            ::::::    ||
10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 26

60 CTCGGGCGTCCTGGTGCATCCCAGTGGGTGCGTCAGCCGCCACACTGTT 109
   :::||||||||||||||||||||:::||||||||:::||||||||||||
26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43

110 TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC 159
    ::::||||  ||||  ||||  ::|||||  |||||:   :||
43 leArgAsnLysSerValIleLeuGlyLeuGlyLeuHisSerLeuPheHisPro 59

160 CAAGAGGCAGGGAGCCAGATGGTGGAGGCCAGCCTCCGTACGGCCACCC 209
    ::::::||  ||||      :::||||:::|||  |||  ||||||
60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75

210 AGAGTACAACAGACCCTTGCTCGCTAAC.................... 237
    : ||| : ||||  ||||
75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92

238 .........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 276
         ||||||||||||||||:::||
92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108

277 TCTGACACCATCCGGAGCATCAGCATTGCTTCCAGTGCCCTACCGCGGG 326
    :::|||:::|||||||||||||||:::||| ||:::|
109 ThrAspAlaValLysValMetAspLeuProThrGlnProAlaLeuGl 125

327 GAACTCTTGCCTCGTTTCTGGCTGGGCTCTGCTGGCCAACGGTGAGCTCA 376
    |:::::|||  ::||||||||||| ||:  :||
125 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu.......... 137

377 CGGGTGTGTGTCTGCCCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGG 426
    ||||                 |||       ||
137 .............................................  137

427 CTGACCCAGAGCTCTGCCTCCCAGGCAGAATGCTACCGTGCTGCAGTGC 476
    ||||||                ||        ||||||||
138 ...ProGluGluPheLeu.......ThrProLysLeuGlnCys 149
```

Percent Similarity: 61.811   Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x PCT-US95-06157-8   .

Align seg 1/1 to: PCT-US95-06157-8 from: 1 to: 238

```
 10 CACTCGGCAGCCCTGGCAGCGGCACTGTCATGGAAAACGAATTGTCTG    59
    ||||||||||||||||||||:::|||:|            :::::      ||
 11 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy   27

 60 CTCGGGCGTCGTCGTCATCCGCAGTGGTGCTGTCAGCCCACACTGTT    109
    :::||||||||||||||||||||||||||||::||||||||||||
 27 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL   44

110 TCCAGAACTCCTACACACCTGGGCTGGCCTGCACAGTCTTGAGGCCGAC   159
    :::::::::        :::  |||||  |||:|||  |||:::||||
 44 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro   60

160 CAAGAGCCAGGGAGCCAGATGTGGAGGCCAGCCTCTCCGTACGGCACCC   209
    :::::::   |||   |||   :::||| ||| |||  |||||
 61 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr   76

210 AGAGTACAACAGACCCTTGCTC............................   231
    |  |||||||  |||||||
 76 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA   93

232 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG   276
    :::::|||||||:::||||||:|||||||||||  |||||||||:||||
 93 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle   109

277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGG   326
    ::||:::|||||||||||  :::::|||  :::|||  |||||||  |
110 ThrAspValValLysValLeuGlyLeuProThrGlnProAlaLeuGl    126

327 GAACTCTTGCCTCGTTTCTGGCTGGGGTCTGCTGGCCGAACGGTGAGCTCA   376
    |::::||||    ::|||||||||||||   :::  |||
126 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu............   138

377 CGGGTGTGTGTCGCCCTCTCAAGGAGGTCCTCTGCCCAGTCGCGGGGG    426
    |||||||
138 ..................................................   138

427 CTGACCCAGAGCTCTGCGTCCAGGCAGCAGAATGCCTACCGTGCTGCAGTGC   476
    |||||             |||    ||||||||  |||||||||
139 .....ProGluGluPheLeu.........ArgProArgSerLeuGlnCys   150

477 GTGAACGTGTCGGTGGTGTCTGAGAGAGGTCAGTAGAGCTCATGACCC    526
    |||:::|||||||:::||||:::||:|||||  :::::     ::::::
151 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrSerGl   167

527 GCTGAACCACCCCAGCATGTTCTGCCGCCGGCGAGGCCAAGACCAGAAGG   576
    ||||||||||||||||||||||   |||||| |||||    ||||
167 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA   184

577 ACTCCTGCAACGGTGACTCTCGGGGGGGCCCCTGATCTGCAACGGGTACTTG   626
    ||::::|||:||||||||||||||||::|||||||||::|||||||
184 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu   200

627 CAGGGCCTTGTGTCTTTCGGAAAAGCCCCGTGTGGCCAAGTTGGCGTGCC   676
    ||||||:::||:::||||||  ||||||   |||||    ||||||:
201 GlnGlyIleThrSerTrpGlyProGlyProCysAlaLeuProGluLysPr   217

677 AGGTGTCTACACCAACCTCTGCAAATTCACTGAGTGATAGAGAAAACCG   726
    ||::::|||||||||||||||     |||  ::    |||  |||
217 oAlaValTyrThrLysValHisTyrArgLysTrpIleLeuLysTyrThrI   234

727 TCCAGGCCAGT 737
    ::  |||:::
234 leAlaAlaAsn 237
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8

Also (rotated column, right side):

```
477 GTGAACGTGTCGGTGGTGTCTGAGGAGGTCTCGAGGAGCTCTATGACCC   526
    ||||:::::   ||:::||||:::||||||:::|||||:::::::::
150 ValGlnLeuHisValIleSerAsnAspValCysAlaGlnValHisProGl   166

527 GCTGTACCACCCCAGCAGTTCTCGCCGCCGGCGAGGCCAAGACCAGAAGG   576
    |||||||||||||||||||||||||
166 nLysValThrLysPheMetLeuCysAlaGlyTrpThrGlyGlyLyss    183

577 ACTCCTGCCAACGGTGACTCTCGGGGGGGCCCCTGATCGTCCAACGGGTACTTG   626
    :::::|||:||||||||||||||||||||||||||:::||||
183 erThrCysSerGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu   199

627 CAGGGCCTTGTGTCTTTCGGAAAAGCCCCGTGTGGCCAAGTTGGCCGTGCC   676
    ||||||:::||:::||||||   ||||:::
200 GlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgPr   216

677 AGGTGTCTACCACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCG   726
    ||:::::|||||||||||||||    :::    |||:|:::
216 oSerLeuTyrThrLysValHisTyrArgLysTrpIleLysAspThrI     233

727 TCCAGGCCAGT 737
    :: |||:::
233 leValAlaAsn 236
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8

alignment_scores:
Quality: 448.50   Length: 254
Ratio: 2.857   Gaps: 5

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-10

```
seq_documentation_block:
; Sequence 10, Application PC/TUS9506157
; GENERAL INFORMATION:
;   APPLICANT:    Mayo Foundation for Medical Education
;                 and Research
;   APPLICANT:    Hybritech Incorporated
;   APPLICANT:    Tindall, Donald J.
;   APPLICANT:    Young, Charles Y.F.
;   APPLICANT:    Saedi, Mohammed S.
;   TITLE OF INVENTION:  Recombinant HK2 Polypeptide
;   NUMBER OF SEQUENCES:  18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Schwegman, Lundberg & Woessner, P.A.
;     STREET:     3500 IDS Center
;     CITY:       Minneapolis
;     STATE:      MN
;     COUNTRY:    USA
;     ZIP:        55402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US95/06157
;     FILING DATE:
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Raasch, Kevin W.
;     REGISTRATION NUMBER:  35,561
;     REFERENCE/DOCKET NUMBER:  150.148WO1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  612-339-0331
;     TELEFAX:    612-339-3061
; INFORMATION FOR SEQ ID NO:  10:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  244 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
PCT-US95-06157-10
```

```
alignment_scores:
             Quality:    448.50          Length:    254
             Ratio:      2.857           Gaps:        5
Percent Similarity:      61.811    Percent Identity:  37.402
```

```
alignment_block:
US-09-030-606-173 x PCT-US95-06157-10

Align seg 1/1  to: PCT-US95-06157-10  from:  1  to: 244

 10 CACTCCCAGCCCTGGCAGGCGGGCACTGGTCATGGAAAACGAATTGTTCTG 59
    ||||||||||||||||||||||:::|| ::::::::  |||
 17 HisSerGlnProTrpGlnValAlaAlaValTyrSerHisGlyTrpAlaHisCy 33

 60 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGGCACACTGTT 109
    |:::||||||||||||||||||||:|||| ||:||||||||||||||||
 33 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 50

110 TCCAGAACTCCTACACCATCGGCCTGGCCTGGCCTGCACAGTCTTGAGGCCGAC 159
    ::::::::::::::   ::||||||||    ||||:::||
 50 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 66

160 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCCGTACGGCCACCC 209
    :::::  ||| ||| ||| ::||||   || ||||    ||||
 67 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 82

210 AGAGTACAACAGACCCTTGCTC.................. 231
    |  ||||| ||||||||||
 82 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 99
```

```
232 ....GCTAACGACCTCATCGTCATCAAGTTGGACGAATCCGTGTCCGAG 276
        ::::::||||||||||::::||::||||:||:|||
 99 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 115

277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGG 326
    :::||||::::::::  :::::::  :::||| ||||::  ||
116 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 132

327 GAACTCTTGCCTCCGTTTCTGGCTGGGGTCTGCTGGCCAACCGTGAGCTCA 376
    |::::||||  :::||||||||||||||   ::: |||
132 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu.......... 144

377 CGGGTGTGTGTCTGCCCTCTTCAAGGAGGTCTCTGCCCAGTCGCGCGGGGG 426
                                                   .....
144 ............................................... 144

427 CTGACCCAGAGCTCGTGCTCCCAGGCAGGAATGCCTACCGTGCTGCAGTGC 476
    :::::||| ||  |||||||||:: |||::||| :||||||||||||||
145 ....ProGluGluPheLeu.......ArgProArgSerLeuGlnCys 156

477 GTGAACGTGTCGGTGGTCTGTGAGGAGGTCTGCAGTAAGCTCTATGACCC 526
    ||:::||| :::::::||::::||||:::||||:|::  |||::: |||:
157 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaAlaTyrSerGl 173

527 GCTGTACCACCCCAGCATGTTCTGCGCCGGCGGAGGCAAGACCAAGGAAGG 576
    |:::::||| ||||||||||||||||||||||||||  |||
173 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA 190

577 ACTCCTGCAACGTGACTCTGGGGGCCCCTGATCTGCAACGGTACTTG 626
    |:::||||:::||||||||||||||||||::|||||||||
190 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 206

627 CAGGGCCTTGTGTTCTTCGGAAAAGCCCCGTGTGGCCAAGTTGGCGTGCC 676
    |||||||::::::||||||||  |||||:::
207 GlnGlyIleThrSerTrpGlyProGluProCysAlaLeuProGluLysPr 223

677 AGGTGTCTACACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCG 726
    |:::||||||||||:::::::  :::|||:: ::::|||||:: :||
223 oAlaValTyrThrLysValValHisTyrArgLysTrpIleLysTyrThrI 240

727 TCCAGGCCAGT 737
    :: ||||::
240 leAlaAlaAsn 243
```

```
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-5

seq_documentation_block:
; Sequence 5, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
;   APPLICANT:   Bandman, Olga
;   APPLICANT:   Goli, Surya K.
;   TITLE OF INVENTION:  A NOVEL PROSTATE-SPECIFIC
;   TITLE OF INVENTION:  KALLIKREIN
;   NUMBER OF SEQUENCES:  5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;     STREET:   3174 Porter Drive
;     CITY:     Palo Alto
;     STATE:    CA
;     COUNTRY:  US
;     ZIP:      94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/744,026
;     FILING DATE:  Herewith
;     CLASSIFICATION:  514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
```

```
        FILING DATE:
        ATTORNEY/AGENT INFORMATION:
;         NAME: Billings, Lucy J
;         REGISTRATION NUMBER: 36,749
;         REFERENCE/DOCKET NUMBER: PF-0154 US
;       TELECOMMUNICATION INFORMATION:
;         TELEPHONE: 415-855-0555
;         TELEFAX: 415-843-4166
;   INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 261 amino acids
;         TYPE: amino acid
;         STRANDEDNESS: single
;         TOPOLOGY: linear
;       IMMEDIATE SOURCE:
;         LIBRARY: GenBank
;         CLONE: 190553
US-08-744-026-5

alignment_scores:
        Quality:  448.50           Length:       254
        Ratio:      2.803          Gaps:           5
Percent Similarity: 62.992     Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x US-08-744-026-5

Align seq 1/1 to: US-08-744-026-5     from: 1 to: 261

 10 CACTCGCCAGCCCTGGCAGGCGGCCACTGGTCATGCATGGAAAACGAATTGTTCTG   59
    |||||||||||||||||||||| :::::::: :::::::   ::|||||||||||
 34 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy         50

 60 CTCGGCGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT     109
    |::|||||||||||||||||||||||||||||::::|||||||||||||
 50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI       67

110 TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC       159
    ::|||::: :::||| ||||| ||||||||  |||||||
 67 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro       83

160 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC        209
    :::::: ||| ||: ||||||||||||||||||||||:::::||||||
 84 GluAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr       99

210 AGAGTACAACAGACACCCTTGCTCGCTAAC................            237
    | ||::: |||||| |||
 99 OLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA      116

238 ........GACCTCAATGCTCATCAAGTTGGACGAAATCCGTGTCCGAG        276
    |||||||||||||:::||| :::::
116 spSerSerHisAspLeuMetLeuArgLeuSerGluSerGluProAlaGluLeu   132

277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCCGGG       326
    ::|||||||||| ::::||| |||:: ||: |||
133 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGi      149

327 GAACTCTTGCCTCCGTTTCTGGCTGGGTCGCTGCTGCGGCAACGGTGAGCTCA   376
    |::::||| ::|||||||||||| ::: |||
149 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu..........      161

377 CGGGTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGG       426

161 ................                                         161

427 CTGACCCAGAGCTCTGCGTCCCAGGCAGCAGAATGCCTACCGTGCTCGCAGTGC  476
    |||||    ::: ||| ||||||||||||||
162 ....ProGluPheLeu.......ThrProLysLeuGlnCys              173

477 GTGAACGTGTCGGTGTGTCTGAGGAGGTCTGCAGTAAGCTCTATGACCC       526
    ||||::: :||||||||:::::||||||::::::
174 ValAspLeuHisValIleSerArgAspValCysAlaGlnValHisProGl      190
```

```
527 GCTGTACCACCCAGCATGTTCTGCCGCCGGCGGCGAGGCCAAGACCAGAAGG     576
    |||::||||||||
190 nLysValThrLysPheMetLeuCysAlaGlyArgTrpThrGlyGlyLyss      207

577 ACTCCTGCAACGGTGACTCTGGGGGCCCCTGATCGTCAACGGGTACTTG       626
    ::::::|||:|||||||||||||||||::::||||||| |||
207 erThrCysSerGlyAspSerGlyArgGlyProLeuValCysAsnGlyValLeu   223

627 CAGGGGCCTGTGTCTTTCGGAAAAGCCCCGTGGCCAAGTGGCGTGCC        676
    |||||:||::::|||||||||||||||:::
224 GlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgPr      240

677 AGGTGTCTACACCAACCTCGAAATTCACTGAGTGATAGAGAAAACCG        726
    |::::||||:::::: ::: :::||||||:::  |||
240 oSerLeuTyrThrLysValHisTyrArgLysTrpIleLysAspThrI        257

727 TCCAGGCCAGT        737
    |||:::
257 leValAlaAsn       260
```

seq_name: /cgn2_6/ptcdata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-6

```
seq_documentation_block:
;  Sequence 6, Application PC/TUS9506157
;  GENERAL INFORMATION:
;       APPLICANT: Mayo Foundation for Medical Education
;       APPLICANT: and Research
;       APPLICANT: Hybritech Incorporated
;       APPLICANT: Tindall, Donald J.
;       APPLICANT: Young, Charles Y.F.
;       APPLICANT: Saedi, Mohammed S.
;       TITLE OF INVENTION: Recombinant HK2 Polypeptide
;       NUMBER OF SEQUENCES: 18
;       CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
;         STREET: 3500 IDS Center
;         CITY: Minneapolis
;         STATE: MN
;         COUNTRY: USA
;         ZIP: 55402
;       COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: PCT/US95/06157
;         FILING DATE:
;         CLASSIFICATION:
;       ATTORNEY/AGENT INFORMATION:
;         NAME: Raasch, Kevin W.
;         REGISTRATION NUMBER: 35,561
;         REFERENCE/DOCKET NUMBER: 150.148WO1
;       TELECOMMUNICATION INFORMATION:
;         TELEPHONE: 612-339-0331
;         TELEFAX: 612-339-3061
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH: 261 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;       MOLECULE TYPE: protein
PCT-US95-06157-6

alignment_scores:
        Quality:  448.50            Length:       254
        Ratio:      2.857           Gaps:           5
Percent Similarity: 61.811      Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x PCT-US95-06157-6
```

APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS
APPLICANT: NO. 5737768ris, Franklin H.
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center/Patent Division
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/361,395
FILING DATE: 22-DEC-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Blalock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X8350A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-361-395-1

alignment_scores:
Quality: 448.00   Length: 244
Ratio: 2.835   Gaps: 5
Percent Similarity: 64.754   Percent Identity: 38.934

alignment_block:
US-09-030-606-173 x US-08-361-395-1   ..

Align seg 1/1   to: US-08-361-395-1 from: 1 to: 244

   13 TCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTC   62
       |||:::|||::||||||||||| :::|::||::||:::||::|:||
   32 SerHisProTyrGlnAlaAlaLeuTyrThrSerGlyHisLeuLeuCysGl   48

   63 GGGGTCCTGGTCGCATCCCAGTGGGTGCTGTCAGCCGCCACACTGTTCC   112
       :|!|||||||:!||||| ||||:::|||:|||||||||||||||
   48 yGlyValLeuHisProLeuTrpValLeuThrAlaAlaHisCysLysL   65

  113 AGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTT.........GAG   153
       : ::: |||: |||||| ||| ::|| :||:::||::||
   65 ysProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArgGlu   81

  154 GCCGACCAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACG   203
       ::::::|||| :::|||||||||:::||:||| || Vall   94
   82 SerSerGlnGluGln...SerSerValArgAla.........Valll   94

  204 GCACCCAGAGTACAACAGACACCCTTGCTCGCTAACGACCTCATGCTCATCA   253
       |||||:::|:||||:::| :::::||| ||:||||||||||||:::
   94 eHisProAspTyrAspAlaAlaSerHisAspIleMetLeuLeuA   111

  254 AGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCAATT   303
       :|||| |:|:::| |||:::|| ||:::|::
  111 rgLeuIleArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu   127

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-361-395-1

seq_documentation_block:
; Sequence 1, Application US/08361395
; Patent No. 5733768
; GENERAL INFORMATION:

Align seg 1/1   to: PCT-US95-06157-6 from: 1 to: 261

   10 CACTCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG   59
       |||||||||||||||||||::|::   ::::||
   34 HisSerGlnProTrpGlnAlaAlaValTyrSerHisGlyTrpAlaHisCy   50

   60 CTCGGGCGTCCTGGTCGCATCCCAGTGGGTGCTGTCAGCCGCCACACTGTT   109
       |:::|||||||||||||||||||||:::|||||:::::|||||||||||||
   50 sGlyGlyValLeuHisProGlnTrpValLeuThrAlaAlaHisCysLysL   67

  110 TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTCGAGGCCGAC   159
       :::::||::: |||||| |||:|||| ||:::||
   67 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro   83

  160 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC   209
       :::::: ||| ||| ::|||| |||:::|| ::|||||
   84 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr   99

  210 AGAGTACAACAGACCCTTGCTC.................   231
       | |||||| ||||:::|
   99 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA   116

  232 ....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG   276
       :::::||||||||||:::|::   :::::
  116 spSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle   132

  277 TCTGACACCATCCGGAGCATCAGCAATTGCTTCGCAGTGCCCTACCGCGGG   326
       ::|||:::||||||||||||||||||||| ||::|:||
  133 ThrAspValIleLysValLeuProThrGlnLeuProAlaLeuLeuGl   149

  327 GAACTCTGCCTCGTTCGGCGGGGTGCTGCTGGCGGAACGGTGAGCTCA   376
       |:::::||| ||||||||||||:::: |||
  149 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu..........   161

  377 CGGGTGGTGTGTCTGCCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGG   426
         161

  161   161

  427 CTGACCCAGAGCTCTGCGTCCCAGGCAGCAGAATGCCTACCGTGCTGCAGTGC   476
       ||||| :: ||| ||||||||||:::
  162 ....ProGluGluPheLeu........ArgProArgSerLeuGlnCys   173

  477 GTGAACGTGTCGGTGGTGGTCTGAGGAGGTCGCAGTAAGCTCCATGAGCCC   526
       |||:::::: :::::::|||||||:::|||::: ||:::
  174 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaAlaArgAlaTyrSerGl   190

  527 GCTGTACCACCCCAGCATGTTCTGCGCCCGGCGGAGGGCAAGACCCAGAAGG   576
       ||||||||||:::::||||||||| ||||
  190 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA   207

  577 ACTCCTGCAACGGTGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTG   626
       ::::||||||||||||||||:::||||||||||| ||
  207 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu   223

  627 CAGGGCCTTGTGTCTTCGGAAAAGCCCCGTGGCCCAGTTGGCCGTGCC   676
       ||||||:::||||:::|||| ||||||
  224 GlnGlyIleThrSerTrpGlyProGluProCysAsnLeuProGluLysPr   240

  677 AGGTGTCTACACCACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCG   726
       |:::::|||||::|| ::::|||||||::: |||
  240 oAlaValTyrThrLysValHisValHisTyrArgLysTrpIleLysTyrThrI   257

  727 TCCAGGCCAGT 737
       |||||::
  257 leAlaAlaAsn 260

```
;         LENGTH: 237 amino acids
;         TYPE:  amino acid
;         STRANDEDNESS:  single
;         TOPOLOGY:  linear
;         MOLECULE TYPE:  peptide
;         IMMEDIATE SOURCE:
;         CLONE:  Amino acid sequence of hK2
;         CLONE:  (deduced from cDNA sequence)
US-08-096-946-10

alignment_scores:
             Quality:   447.50             Length:     254
               Ratio:    2.850               Gaps:       5
   Percent Similarity:  61.811   Percent Identity:  37.402

alignment_block:
US-09-030-606-173 x US-08-096-946-10

Align seg 1/1  to: US-08-096-946-10  from: 1  to: 237

  10 CACTCGCAGCCTGGCAGGCCGCACTGGTCATGAGAAAACGAATTGTTCTG  59
     |||||||||||||||||||||||||:::|||   :::::    ||||||
  10 HisSerGlnProTrpGlnValAlaValTrpSerHisGlyTrpTrpAlaHisCy  26

  60 CTCGGGCGTCCTGCATCCGCAGTGGTGTCTGTCAGCCGCCACACTGTT 109
     |:::|||||||||||||||||||||||||||:|||||||||||||:::
  26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL  43

 110 TCCAGAACTCCTACACCATCGGCGTGGCCTGCACAGTCTTGAGGCCGAC 159
     ::::::::   ||||||   :::|||   ||||:::|||
  43 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro  59

 160 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 209
     :::::|||   |||   |||   |||||:::|||   |||||   ||||
  60 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr  75

 210 AGAGTACCAACAGACCCCTGCTC..........................  231
     |||||   |||||
  75 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA  92

 232 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 276
          |||||:::|||||||||:::|||   |||     :::|||
  92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 108

 277 TCTGACACCATCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGGCGGG 326
     :::|||||||||   :::|||   |||:::   |||||||||||||||||
 109 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 125

 327 GAACTCTTGCCTCGTTTCTGGCTGGGGTCTCGCTGGCGAACGGTGAGCTCA 376
     |||||:::|||   |||||||||||:::   :::|||
 125 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu...........  137

 377 CGGGTGTGTGTCTGCGCCCTCTTCAAGGAGGGTCCTCGCCCAGTCGCGGGGG 426
     ..........
 137 ..........................................  137

 427 CTGACCCAGAGCTCGTCGTCCCAGGCAGGCCAGAATGCCTACCGTGCTGCAGTGC 476
     |||||     :::    |||   |||||||||   |||||||||
 138 .....ProGluGluPheLeu........ArgProArgSerLeuGlnCys 149

 477 GTGAACGGTCGGTGGTGTCTGAGGAGGTCTCCAGTAAGCTCTATGACCC 526
     |||:::   :::||||||   :::|||||:::   |||   :::
 150 ValSerLeuHisLeuSerAsnAspMetCysAlaArgAlaTyrSerGl 166

 527 GCTGTACCACCCCAGCATGTTCTGCGCCGGCAAGACCAGGAGG 576
     |||||||||||||||||:::||||||   |||   ||||||
 166 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyLyLysA 183

 577 ACTCCTGCAACGGTACTGACTCTGGGGGCGCCCCTGATCTGCAACGGGTACTTG 626
     |||||||||||||||:::||||||||||:::||||||
 183 spThrCysGlyCysGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199
```

```
 304 GCTTCGCAGTGCCCTACCGGCGGGGAACTCTTGCCTCCGTTTCTGGCTGGGG 353
     :::|||   :::                ||||   |||||||
 128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTrpGl 144

 354 TCTGCTGGCGAACGGTGAGCTCACGGGTGTGTGCCCTCTTCAAGGA 403
     |
 144 y.............................................  144

 404 GGTCCTCTGCCCAGTCGCGGGGGGTGACCCAGAGCTCTGCGTCCCAGGCA 453
     ::::::|||    |||
 145 ..................LysThrAlaAsp..GlyA 150

 454 GAATGCCTACCGTGCTGCAGTGCGTGAACGGTCGGTGGTGTCTGAGGAG 503
     :::|||   :::|||||:::   :::|||||:::
 150 spPheProAspThrIleGlnCysAlaTyrIleHisLeuValSerArgGlu 166

 504 GTCTCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATGTTCTGCGC 553
     |||:::   :::||||
 167 GluCysGluHisAlaTyrProGlyIleThrGlnIleThrGlnMetLeuCysAl 183

 554 CGGCGGAGGGCAAGACCAGGAGGACTCCTGCAACGGTGACTCTGGGGGC 603
     |||:::   |||||||||||||||
 183 aGlyAspGluLysTyrGlyLysAspSerCysGlnGlyAspSerGlyGlyP 200

 604 CCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTCGTCTTTCGGAAAGCC 653
     |||||:::   :::|||||   :::|||   |||:::|||||
 200 roLeuValCysGlyAsnHisLeuArgGlyLeuValSerTrpGlyAsnIle 216

 654 CCGTGTGGCCAAGTTGGCGTGCCCAGTGTCTACACCAACCTCTGCAAATT 703
     |||||||||||||||||||||||||   :::|||||
 217 ProCysGlySerLysGluLysProGlyValTyrThrAsnValCysSerArgTy 233

 704 CACTGAGTGGATAGAAGAAACCGTCCAGGCC 734
     :||||||||||:::||||||:::||||||
 233 rThrAsnTrpIleGlnLysThrIleGlnAla 243
```

```
 60 CTCGGGCCTCTCGGTGCATCCCCAGTGGGTGGTGTGTCAGCCGCACACTGTT 109
    |||:||||||||||||||||||||||||||||||||:::|||||||||||
 26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 43

110 TCCAGAACTCCTACACCATCGGGCTGCGGGCTGGCGCACAGTCTTGAGGCCGAC 159
    ::::::::  ::  :::  |||||| |||:|||
 43 euLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 59

160 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCCTCTTCCGTACGGCACCC 209
    :::::  ||  |||  :::|||  |||:||  ||||
 60 GluAspThrGly..GlnArgValProValSerHisSerPheProHisPr 75

210 AGAGTACAACAGACCCTTGCTC.................... 231
    |  |||||  |||||
 75 oLeuTyrAsnMetSerLeuLysHisGlnSerLeuArgProAspGluA 92

232 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCCAG 276
    :::::||||||||||||:::||:::||  :::::
 92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 108

277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGG 326
    :||::::||||||||  :::|||  ::||||  |||::::|
109 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 125

327 GAACTCTTGCCTCGTTCTGGCTGGGGTCCTGCTGGCGAACGGTGAGCTCA 376
    ||:::|||  :::|||||||||||| :: ||
125 yThrThrCysTyrAlaSerGlyTrpGlyLysSerIle..Glu......... 137

377 CGGGTGTGTGTCTCCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGG 426
    ........................
137 ................ 137

427 CTGACCCAGAGCTCTGCGTCCCAGGCAGCAGAATGCCTACCGTGTGCAGTGC 476
    |||||  ||  :::
138 .....ProGluGluPheLeu.......ArgProArgSerLeuGlnCys 149

477 GTGAACGTCTCGGTGGTCGTCTGAGGAGGTCGAGTAAGCTCTATGACCC 526
    ||:::||||||:::||:::::  :::|||:::  |||:::||
150 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrSerGl 166

527 GCTGTACCACCCCAGCATGTTCTCGCCGCGGAGGGCCAAGACCAGAAGG 576
    ||:::||||||||  |||||||||  |||
166 uLysValThrGluMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA 183

577 ACTCCTGCAACGGTGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTG 626
    ||:::||:::||||||||||||||:::||||||  :::||||
183 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199

627 CAGGGCCTTGTGTCTTTCGGAAAAGCCCGTGTGGCCAAGTTGGCGTGCC 676
    |||||||||  |||||||  |||
200 GlnGlyIleThrSerTrpGlyProGluProCysAlaLeuProGluLysPr 216

677 AGGTGTCTACACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCG 726
    |:::||||||||||:::  :::||||||::  |||:
216 oAlaValTyrThrLysValValHisTyrArgLysTrpIleLysAspThrI 233

727 TCCAGGCCAGT 737
    ::|||:::
233 IeAlaAlaAsn 236
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-16

seq_documentation_block:
; Sequence 16, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18

```
627 CAGGGCCTTGTGTCTTTCGGAAAAGCCCGTGTGGCCAAGTTGGCGTGCC 676
    |||||||::||||||:::||||  |||||||::  ||
200 GlnGlyIleThrSerTrpGlyProGluProCysAlaLeuProGluLysPr 216

677 AGGTGTCTACACCAACCTCGCAAATTCACTGAGTGGATAGAGAAAACCG 726
    |:::||||||||:::::::  ::  ::||||||::  |||:
216 oAlaValTyrThrLysValValHisTyrArgLysTrpIleLysAspThrI 233

727 TCCAGGCCAGT 737
    :: ||:::: ||
233 IeAlaAlaAsn 236
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-10

seq_documentation_block:
; Sequence 10, Application PC/TUS9407329
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical
; Education and Research
; TITLE OF INVENTION: Antibodies Specific for Human
; TITLE OF INVENTION: Prostate Glandular Kallikrein
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner
; STREET: 3500 IDS Center
; STREET: 80 South Eighth Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07329
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,651
; REFERENCE/DOCKET NUMBER: 150.62WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; CLONE: Amino acid sequence of hK2
; CLONE: (deduced from cDNA sequence)
PCT-US94-07329-10

alignment_scores:
    Quality: 447.50   Length: 254
    Ratio: 2.850   Gaps: 5
Percent Similarity: 61.811   Percent Identity: 37.402

alignment_block:
US-09-030-606-173 x PCT-US94-07329-10

Align seg 1/1 to: PCT-US94-07329-10 from: 1 to: 237

```
 10 CACTCGGCCAGCCCTGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTG 59
    |||||||||||||||:::||:::  ::  :::  |
 10 HisSerGlnProTrpGlnValAlaValTrpSerHisGlyTrpAlaHisCy 26
```

```
CORRESPONDENCE ADDRESS:
  ADDRESSEE:  Schwegman, Lundberg & Woessner, P.A.
  STREET:  3500 IDS Center
  CITY:  Minneapolis
  STATE:  MN
  COUNTRY:  USA
  ZIP:  55402
COMPUTER READABLE FORM:
  MEDIUM TYPE:  Floppy disk
  COMPUTER:  IBM PC Compatible
  OPERATING SYSTEM:  PC-DOS/MS-DOS
  SOFTWARE:  PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER:  PCT/US95/06157
  FILING DATE:
  CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
  NAME:  Raasch, Kevin W.
  REGISTRATION NUMBER:  35,561
  REFERENCE/DOCKET NUMBER:  150.148WO1
TELECOMMUNICATION INFORMATION:
  TELEPHONE:  612-339-0331
  TELEFAX:  612-339-3061
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
  LENGTH:  237 amino acids
  TYPE:  amino acid
  STRANDEDNESS:  single
  TOPOLOGY:  linear
  MOLECULE TYPE:  peptide
PCT-US95-06157-16
```

```
alignment_scores:
        Quality:  447.50          Length:      254
        Ratio:    2.850           Gaps:          5
Percent Similarity:  61.811     Percent Identity:  37.402
```

```
alignment_block:
US-09-030-606-173 x PCT-US95-06157-16

Align seg 1/1  to: PCT-US95-06157-16   from: 1  to: 237

  10 CACTCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG  59
     |||||||||||||||||||::|||::|||         ::::::::  ||
  10 HisSerGlnProTrpGlnValAlaValTrpSerHisGlyTyrAlaHisCy  26

  60 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 109
     |:::||||||||||||::|||||||||||||:::||||||||||||||:
  26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL  43

 110 TCCAGAACTCCTACACCATCGGGCTGGGCCTGGCCTGCACAGTCTTGAGGCCGAC 159
     ::::::::::   ::  ||||||  |||::|||       ||::||
  43 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro  59

 160 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCCTCCGTACGGCACCC 209
     ::::::|||   ||||| :: |||  ::||| |||  |||||    ||||
  60 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr  75

 210 AGAGTACAACAGACCCTTGCTC.................... 231
     ||||||    |||||
  75 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA  92

 232 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 276
     ::::::|||||||||||||::::||||   :::||||||||||||||:
  92 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 108

 277 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGG 326
     ::::||||||::|||||::||||||      ::|||   ||||::|| ||
 109 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuGl 125

 327 GAACTCTTGCCTCGTTTCTGGCTGGGGTCTCCTGGCCAACCGGTGAGCTCA 376
     |:::::::|||   ::||||||||||||||||      ::: |||
```

```
125 yThrThrCysTyrAlaSerGlyTrpGlySerIle..Glu.......... 137

377 CGGGTGTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGG 426
    ||||||
137 .................................................. 137

427 CTGACCCAGAGCTCGTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGC 476
    ||||||                      |||        ||||||||||
138 .....ProGluGluPheLeu.......ArgProArgSerLeuGlnCys 149

477 GTGAACGTGTCGGTGGTGTCTGAGGAGGTCTCCAGTAAGCTCTATGACCC 526
    |||::::::::   ::::::|||:::||::::::|||   ::::::
150 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrSerGl 166

527 GCTGTACCACCCCAGCATGTTCTGCCCCGGCGGAGGCAAGACCAGAAGG 576
    :::||||||     |||||||||||
166 uLysValThrGluPheMetLeuCysAlaGlyLeuTrpThrGlyGlyLysA 183

577 ACTCCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGTACTTG 626
    |||::||::|||||||||||||||||||::|||||||||||||  ||
183 spThrCysGlyGlyAspSerGlyGlyProLeuValCysAsnGlyValLeu 199

627 CAGGGCCTTGTGTCTTTCGGAAAAGCCCCGTGTGGCCAAGTTGGCGTGCC 676
    ||||:::::::|||:::||    ||||||  |||||::
200 GlnGlyIleThrSerTrpGlyProGluProCysAlaLeuProLysPr 216

677 AGGTGCTACACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCG 726
    :::||||||||::::::::   :::|||||||:::   ::||||:::  |||
216 oAlaValTyrThrLysValHisTyrArgLysTrpIleLysAspThrI 233

727 TCCAGGCCAGT 737
    :: :||:::
233 leAlaAlaAsn 236
```

us-09-030-606-173.rai

Wed Sep 29 14:27:13 1999

OM of: US-09-030-606-174 to: Issued_Patents_AA:*　　out_format : pfs

Date: Sep 25, 1999　4:01 PM

About: Results were produced by the GenCore software, version 4.5,
　　　　Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/USPTO_spool/US09030606/runat_24091999_171616_29813/app_query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEX=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-174
Query length: 1459
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1

alignment_scores:
　　　　Quality:　623.00　　　　　　Length:　187
　　　　Ratio:　4.357　　　　　　Gaps:　4
　　Percent Similarity:　76.471　　Percent Identity:　70.053

alignment_block:
US-09-030-606-174 x US-08-744-026-1

Align seg 1/1　to: US-08-744-026-1　from: 1　to: 248

```
     1 TCAGCCGCACACTGTTTCCAGAAGTGAGTGCAGAGTCCTCACCACCATCGG  52
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
    62 SerAlaAlaHisCysPhe.........GlnAsnSerTyrThrIleGl  74
                               |||::||:||||||||||||||
    53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCCAGAGGCCAGGATGG 102
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
    74 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV  91
                            |||||||||||||||||||||||
   103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1 +		623.00	1175.52	8.8e-59	248	!
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-17111A-121 -		292.50	533.42	3.2e-23	3	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-454-557C-121 -		268.00	486.89	1.3e-20	375	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-278-091-9 +		226.50	413.03	2.9e-16	223	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-9 +		226.50	413.03	2.9e-16	223	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9 +		226.50	413.03	2.9e-16	223	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-9 +		226.50	413.03	2.9e-16	223	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2 +		222.00	402.88	9.4e-16	253	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2 +		222.00	402.88	9.4e-16	253	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-546-12-1 +		218.00	396.55	2.4e-15	225	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-9 +		210.50	379.30	1.7e-14	286	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-9 +		210.50	379.30	1.7e-14	286	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-9 +		210.50	379.30	1.7e-14	286	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-10 +		191.50	344.90	1.7e-12	237	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-10 +		191.50	344.90	1.7e-12	237	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-16 +		191.50	344.90	1.7e-12	237	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1 -		191.50	344.85	1.7e-12	238	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8 +		191.50	344.56	1.7e-12	244	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-10 +		191.50	343.77	1.8e-12	261	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-10 +		191.50	343.33	1.8e-12	271	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-10 +		191.50	343.33	1.8e-12	271	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-10 +		191.50	343.33	1.8e-12	271	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-3 +		189.50	340.05	2.9e-12	258	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-7 +		188.50	339.57	3.5e-12	228	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-7 +		188.50	339.57	3.5e-12	228	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-09303-1 +		187.00	336.08	5.2e-12	240	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-7 +		188.50	339.57	3.5e-12	228	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-7 +		188.50	339.57	3.5e-12	228	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11 +		188.50	339.12	3.6e-12	237	!
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-1 +		188.50	339.12	3.6e-12	237	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-790-137-4 +		187.50	337.90	3.8e-12	263	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-7 +		187.50	336.07	4.8e-12	261	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-5 +		187.50	336.07	4.8e-12	261	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-472-228A-1 +		187.00	336.08	5.2e-12	240	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-4 +		182.50	326.39	1.7e-11	262	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-1 +		182.50	326.39	1.7e-11	262	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-3 +		182.50	326.39	1.7e-11	262	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-681-151-4 +		182.50	326.39	1.7e-11	262	
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-7 +		182.50	325.57	1.7e-11	281	
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-7 +		182.50	325.57	1.7e-11	281	
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-7 +		182.50	325.57	1.7e-11	281	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-8 +		182.50	324.85	1.8e-11	299	!

/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-8 + 182.50 324.85 1.8e-11 299 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-8 + 182.50 324.85 1.8e-11 2
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-844-024-2 + 181.50 325.64 2.0e-11 237 !

```
 91 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu 107
153 GCTAAGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
    |||||||||||||||||||||||||||||||||||||||||||||||
108 AlaAsnAspLeuMet***IleLysLeuAspGluSerValSerGluSerAs 124
203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAACT 252
    |:::|||||||:::||||||||:::|||||||||||||||||||||||
124 pAsnIleArg***IleSerIle***SerGlnCysProThrAlaGlyAsnP 141
253 CTTGCCTCGTTTCTCGGCTGGGGTCTCGCTGGCGAACGGTGAGCTCACGGGT 302
    |||||||||||||||||||||||||||
141 heCysLeuValSerGlyTrpGlyLeuLeu.............. 150
303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTTCTGCCCAGTCGCGGGGGCTGAC 352
150 ................ 150
353 CCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
    ||||:  |||||||||||||||||||||||||||
151 .........AlaAsn..GlyArgMetProThrValLeuGlnCysValAs 163
403 CGTGTCGGTGGTGTCTGANGAGGTCTGCANTAAGCTCTATGACCCGGTGT 452
    |||||||||| || :::||||||||||||||||||||||
163 nValSerValValSerGluGluValCysSerLysLeuTyrAspProLeuT 180
453 ACCACCCCANCATGTTCTGCGCCGCGGAGGGCAAGACCAGAAGGACTCC 502
    |||||||::||||||||||||||||||||||||||||||||||||
180 yrHisProSerMetPheCysAlaAlaGlyLyGlyGlnAsnGlnLysAspSer 196
503 TGCAACGTGAGAGAGGGGAAAGGGGAGGGAGGCCAGGCGACTCAGGAAGGTG 552
    |||||  |||  |||  :::||||:
197 CysAsn.........GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 210
553 GAGAGGGGG 562
    :  :::|||
210 rLeuGlnGly 213
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-17111A-121

seq_documentation_block:
; Sequence 121, Application PC/TUS9517111A
; GENERAL INFORMATION:
;   APPLICANT: de la Monte, Suzanne
;   APPLICANT: Wands, Jack R.
;   TITLE OF INVENTION: Neural Thread Protein Gene Expression and
;   TITLE OF INVENTION: Detection of Alzheimer's Disease
;   NUMBER OF SEQUENCES: 121
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
;   STREET: 1100 New York Avenue, Suite 600
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/17111A
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/340,426
;   FILING DATE: 14-NOV-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ludwig, Steven R.
;   REGISTRATION NUMBER: 36,203
;   REFERENCE/DOCKET NUMBER: 0609.3840002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600

      TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 121:
   SEQUENCE CHARACTERISTICS:
     LENGTH: 397 amino acids
     TYPE: amino acid
     TOPOLOGY: linear
   MOLECULE TYPE: protein
PCT-US95-17111A-121

alignment_scores:
          Quality: 292.50   Length: 100
          Ratio: 3.611   Gaps: 2
Percent Similarity: 81.000   Percent Identity: 66.000

alignment_block:
US-09-030-606-174/rev x PCT-US95-17111A-121
Align seq 1/1 to: PCT-US95-17111A-121   from: 1 to: 397

```
1459 TTTTTTTTTTTTTTTTGAGACAGAGTCTTACTCTGTTGCCCAGCTG 1410
     ||| |||  ::||||| |||||::||||||||||||||||
 298 PheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGl 314
1409 AGTATAGTGGTGTGATCTCAACTCACTTCAACCTCTGCCTCCCATATTCA 1360
     |||| |||  :::||||||||||||||||||||||  :::|
 314 yValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuL 331
1359 AGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCTG 1310
     |:::|||||||||||||||||||||||||||||||||||  |||
 331 ysArgPheSerCysSerLeuSerLeuProSerSerTrpAspTyr.GlyHisLe 347
1309 CCACCATATCCAGCTAACTTTTGTATTTTAGTACAGACAGGATTCACC 1260
     |||||  ::::||||||||||||||
 347 uHisHisThrProLeuIlePheValPheSerLeuGluAlaGlyPheHisH 364
1259 ATTTTGGCCAGGCTGGTCTGAACTCCTACCTCAAGTGATCTGCCTGCC 1210
     ||::: |||||||||||||||||||||:::||| |||||||
 364 isIleCysGlnAlaGlyLeuLeuLysLeuLeuThrSerGlyValThrProHisProAla 380
1209 TCGGCCTCC.CAAAGTGCTGGGATTACAGGCATGAGC...CACCCTGCC 1166
     ||| ||||||||||||||||||||||::::::| ||||||
 381 SerAlaPheGlnSerAlaGlyIleThrGlyValThrProHisProAla 396
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-454-557C-121

seq_documentation_block:
; Sequence 121, Application US/08454557C
; Patent No. 5830670
; GENERAL INFORMATION:
;   APPLICANT: de la Monte, Suzanne
;   APPLICANT: Wands, Jack R.
;   TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection
;   TITLE OF INVENTION: of Alzheimer's Disease
;   NUMBER OF SEQUENCES: 121
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
;   STREET: 1100 New York Avenue, Suite 600
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: U.S.A.
;   ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/454,557C
;   FILING DATE: 30-MAY-1995
;   CLASSIFICATION: 514
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ludwig, Steven R.

```
;       REGISTRATION NUMBER:   36,203
;       REFERENCE/DOCKET NUMBER:   0609.3840003
;       TELECOMMUNICATION INFORMATION:
;       TELEPHONE:   (202) 371-2600
;       TELEFAX:   (202) 371-2540
;       INFORMATION FOR SEQ ID NO:   121:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH:   375 amino acids
;       TYPE:   amino acid
;       TOPOLOGY:   linear
;       MOLECULE TYPE:   protein
;   US-08-454-557C-121

alignment_scores:
           Quality:   268.00          Length:   79
           Ratio:   4.254             Gaps:   0
Percent Similarity:   79.747     Percent Identity:   64.557

alignment_block:
US-09-030-606-174/rev x US-08-454-557C-121

Align seg 1/1  to: US-08-454-557C-121  from: 1  to: 375

1459 TTTTTTTTTTTTTTTTGAGACAGAGTCTTACTCGTGTCCCCAGCTGG 1410
     ||| ||| |||::|||||| ||||||::||||||::: ||||
297 PheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGlI 313

1409 AGTATAGTGGTGTGATCTCAACTCACTTCAACCTCTGCCTCCATATTCA 1360
     |||  ||| ::|||:::||||||||||||||||||||  :::|
313 yValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuL 330

1359 AGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGATTACAGGGCGCCTG 1310
     ||:::|||||||||||||||||||||||||||||||||| ::::||
330 ysArgPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeu 346

1309 CCACCATATCCAGCTAACTTTTGTATTTTAGTACAGACAGGATTTCACC 1260
     ||||||::|||:|||||||||||||||||  ::||||
347 ProProHisProAlaAsnPheCysIlePheIleArgGlyGlyValSerPr 363

1259 ATTTTGGCCAGCTGTCTTGAACTCCTTACCTCAAG 1223
     |::: |||||||| ||||| |||::: ||||||
363 oTyrLeuSerGlyTrpSerGlnThrProAspLeuArg 375

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB pep:US-08-278-091-9

seq_documentation_block:
;   Sequence 9, Application US/08278091
;   Patent No. 5506139
;   GENERAL INFORMATION:
;       APPLICANT:   LOOSMORE, Sheena M
;       APPLICANT:   YANG, Yan-Ping
;       APPLICANT:   CHONG, Pele
;       APPLICANT:   OOMEN, Raymond P.
;       APPLICANT:   KLEIN, Michel H.
;       TITLE OF INVENTION:   Analog of Haemophilus Hin47 Protein with
;       TITLE OF INVENTION:   Reduced Protease Activity
;       NUMBER OF SEQUENCES:   23
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE:   Sim & McBurney
;           STREET:   Suite 701, 330 University Avenue
;           CITY:   Toronto
;           STATE:   Ontario
;           COUNTRY:   Canada
;           ZIP:   M5G 1R7
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE:   Floppy disk
;           COMPUTER:   IBM PC compatible
;           OPERATING SYSTEM:   PC-DOS/MS-DOS
;           SOFTWARE:   PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER:   US/08/278,091
;           FILING DATE:   21-JUL-1994
```

```
;       CLASSIFICATION:   435
;       ATTORNEY/AGENT INFORMATION:
;       NAME:   Stewart, Michael I
;       REGISTRATION NUMBER:   24,973
;       REFERENCE/DOCKET NUMBER:   1038-371
;       TELECOMMUNICATION INFORMATION:
;       TELEPHONE:   (416) 595-1155
;       TELEFAX:   (416) 595-1163
;   INFORMATION FOR SEQ ID NO:   9:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH:   223 amino acids
;       TYPE:   amino acid
;       STRANDEDNESS:   single
;       TOPOLOGY:   linear
;   US-08-278-091-9

alignment_scores:
           Quality:   226.50          Length:   184
           Ratio:   2.137             Gaps:   6
Percent Similarity:   57.609     Percent Identity:   33.152

alignment_block:
US-09-030-606-174 x US-08-278-091-9

Align seg 1/1  to: US-08-278-091-9  from: 1  to: 223

3 TCAGCCGCCACACTGTTTCCAGAAGTGAGTGAGTGCAGAGCTCCTACCACCATCGG 52
  |!|||||||||||:||||||||||||||:: :::|||
37 SerAlaAlaHisCysThrLysSerGlyIleGln..........ValAr 49

53 GCTGGGCCTGCACAGTCTTGAGCGCGACCAAGAGCCAGGAGCCAGATGG 102
   ||||||       :::::::::: ||| :::::::||:::||       ||
49 gLeuGlyGlusApAsnIleAsnVal...ValGluGlyAsnGluGlnPheI 65

103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
    ::::||||||       |||||::::||||||       ||||||
65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu 81

153 GCTAACGACCTCATCAGTTGGACGAATCCGTGTCCCAGTCTGA 202
    ||||||||||||||||||||       :::||||       ::||||||
82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe 98

203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACT 252
    :: ::        ||||||||||       :::::||| ::::||||||::::
98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115

253 CTTGCCTCGTTTCTGCGTGGGTCTCGCTGGCGAACGGTGAGCTCACGGGT 302
    ::|||||:::||||||||||     ||| :::||       ::::||
115 lnCysLeuIleSerGlyTrpGly.........AsnThrLysSerSerGly 128

303 GTGTGTCGCCCTCTTCAAGGAGGTCCTCGCCCAGTCGCGGGGGCTGAC 352
    ::|||                     |||||||||||       |||||||||
129 ..ThrSer.......................................... 130

353 CCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCGTGAA 402
    |||        |||||||||:::|||:::
131 ...............................TyrProAspValLeuLysCysLeuLy 139

403 CGTGTCGGTGGTGTCTGANGAGGTCTGCANTAAGCTCTATGACCCGCTGT 452
    ::|||||:::||| |||       ::: |||
139 sAlaProIleLeuSerAspSerCysLysSerAlaTyrProGlyGlnI 156

453 ACCACCCANCATGTTCTGCCGCCGGAGGGCAAGACCAGAAGGACTCC 502
    ||||||||||       |||||||||
156 leThrSerAsnMetPheCysAlaGlyTyrLeuGluGlyLysLysAspSer 172

503 TGGCAAC...............GTGAGAGAGGGGAAAGGGAGGGG 531
    |||:::::                       ||||||||||
173 CysGlnGlyAspSerGlyGlyProValValCysSerGlyLysLeuGlnGl 189

532 C 532
```

189 y 189

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-9

```
;  seq_documentation_block:
;  Sequence 9, Application US/08483859
;  Patent No. 5655436
;  GENERAL INFORMATION:
;      APPLICANT:  LOOSMORE, Sheena M.
;      APPLICANT:  YANG, Yan-Ping
;      APPLICANT:  CHONG, Pele
;      APPLICANT:  OOMEN, Raymond P.
;      APPLICANT:  KLEIN, Michel H.
;      TITLE OF INVENTION:  Analog of Haemophilus Hin47 Protein with
;      TITLE OF INVENTION:  Reduced Protease Activity
;      NUMBER OF SEQUENCES:  23
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,859
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,149
;      FILING DATE:  26-AUG-1994
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/278,091
;      FILING DATE:  21-JUL-1994
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I.
;      REGISTRATION NUMBER:  24,973
;      REFERENCE/DOCKET NUMBER:  1038-495 MIS:vg
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (416) 595-1155
;      TELEFAX:  (416) 595-1163
;  INFORMATION FOR SEQ ID NO:  9:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  223 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-483-859-9
```

```
alignment_scores:
    Quality:          226.50          Length:         184
    Ratio:              2.137         Gaps:             6
Percent Similarity:    57.609    Percent Identity:  33.152

alignment_block:
US-09-030-606-174 x US-08-483-859-9

Align seg 1/1  to: US-08-483-859-9  from: 1 to: 223

    3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG  52
      ||||||||||||||||||||||||   ::||||                  :::
   37 SerAlaAlaHisCysTyrLysSerGlyIleGln............ValAr  49

   53 GCTGGGCGCTGCACAGTCTTGAGGCGCCACCAAGAGCCAGGAGCCAGATGG 102
      ||||||    :::::::::||       ||::::||:|||::|||:||:::
   49 gLeuGlyGluAspAsnIleAsnVal...ValGluGlyAsnGluGlnPheI  65

 -103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCCTTGCTC 152
```

```
65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu  81
   ::::::||||| |||::::  |||||::::||||||
153 GCTACGACCTCATGTCATCAAGTTGGACGAANTCCGTGTCCGAGTCTGA 202
    ||||||||||||||| :::::::|||| ::::::|||
82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe  98
    ::: |||||||||| ::::::|| ::||||||::||
203 CACCATCCGGAGCAGCAGCATTGCTTCGCAGTGCCTACCGCGGGGAACT 252
   : ::: |||||||::: :::::|| :::||||||:||
98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115
   ::||||||:::||||||||||  ||  ::: ::|||
253 CTTGCCTCGTTTCTGGCTGGGGTCTGCTGGCGAACGGTGAGCTCACGGGT 302
   ::||||||:::||||||||||    ||| ::: ::||
115 lnCysLeuIleSerGlyTrpGly........AsnThrLysSerSerGly 128
   :::||
303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGAC 352
   ::::||
129 ..ThrSer.................................... 130
   ::||
353 CCAGAGCTCTGCGTCCCAGGCAGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
      ||| ||||||:|||:::::|
131 .................TyrProAspValLeuLysCysLeuLy 139
403 CGTGTCGGTGGTGTCTGANGAGGTCTGCANTAAGCTCTATGACCCGCTGT 452
   ::: ::::::||  ::: ||  |||||
139 sAlaProIleLeuSerAspSerSerCysLysSerAlaTyrProGlyGlnI 156
   ||||||||
453 ACCACCCCANCATGTTCTGCGCCGCGGAGGCAAGACCAGAAGGACTCC 502
      |||||||||
156 leThrSerAsnMetPheCysAlaGlyTyrLeuGluGlyLyGlyLysAspSer 172
   ::::|| |||||||
503 TGCAAC............GTGAGAGAGGGGAAAAGGGGAGGG 531
   ||::::                                  ::||||||   ::||
173 CysGlnGlyAspSerGlyGlyProValCysSerGlyLyLysLeuGlnGl 189
532 C 532
   |
189 y 189
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9

```
;  seq_documentation_block:
;  Sequence 9, Application US/08472173
;  Patent No. 5665353
;  GENERAL INFORMATION:
;      APPLICANT:  LOOSMORE, Sheena M
;      APPLICANT:  YANG, Yan-Ping
;      APPLICANT:  CHONG, Pele
;      APPLICANT:  OOMEN, Raymond P.
;      APPLICANT:  KLEIN, Michel H.
;      TITLE OF INVENTION:  Analog of Haemophilus Hin47 Protein with
;      TITLE OF INVENTION:  Reduced Protease Activity
;      NUMBER OF SEQUENCES:  23
;      CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;      COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,173
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,149
;      FILING DATE:  26-AUG-1994
;      PRIOR APPLICATION DATA:
```

;   APPLICATION NUMBER: US 08/278,091
;   FILING DATE: 21-JUL-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Stewart, Michael I.
;   REGISTRATION NUMBER: 24,973
;   REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (416) 595-1155
;   TELEFAX: (416) 595-1163
;   INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 223 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-472-173-9

alignment_scores:
        Quality: 226.50   Length: 184
        Ratio: 2.137   Gaps: 6
Percent Similarity: 57.609 . Percent Identity: 33.152

alignment_block:
US-09-030-606-174 x US-08-472-173-9

Align seg 1/1 to: US-08-472-173-9 from: 1 to: 223

   3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG 52
     ||||||||||||||||:::::::::: ::||| ::
  37 SerAlaAlaHisCysTyrLysSerGlyIleGln..........ValAr 49

  53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
     ||||| ||| :::||:::|||:::   :
  49 gLeuGlyGluAspAsnIleAsnVal...ValGluGlyAsnGluGlnPheI 65

 103 TGGAGGCCAGCCTCCGTACGGCACCCAGAGTACAACAGACCCCTTGCTC 152
     :::::||||| |||::: |||||:::|||||||||
  65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu 81

 153 GCTAACGACCTCATGCTCATCAGTTGGACGAATCCGTCCGAGTCTGA 202
     ||||:::::||||||||||||| :::::::||| :::::|
  82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe 98

 203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGGAACT 252
     ||::||||||||||||||||||||| :::||:|||||||||||:::::||
  98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115

 253 CTTGCCTCGTTCTGGCTGGGGTCTGCTGCTGGCGAACGGTGAGCTCACGGGT 302
     :::||||||:::|||||||| ||| :::|||
 115 lnCysLeuIleSerGlyTrpGly.......AsnThrLysSerSerGly 128

 303 GTGTGTCTGCCCCTCTTCAAGGAGGTCCTCGCCCAGTCGCGGGGGCTGAC 352
     :::|||
 129 ..ThrSer.....................................   130

 353 CCAGAGCTCTGCGTCCCAGGCAGCAGAATGCTACCGTGCTGCTGCGTGAA 402
     ||| |||:::::|||:::|
 131 ....TyrProAspValLeuLysCysLeuLy 139

 403 CGTGTCGGTGGTCTGTGANGAGGTCTGCANTAAGCTCTATGACCCGCTGT 452
     :::   :::::||| :: ||| ||
 139 sAlaProIleLeuSerAspSerSerCysLysSerAlaTyrProGlyGlnI 156

 453 ACCACCCCANCATGTTCTGCGCCGGCGAGGGCCAAGACCAGAAGGACTCC 502
     ||||||||| ::: ||||||||
 156 leThrSerAsnMetPheCysAlaGlyTyrLeuGluGlyLyGlyLysAspSer 172

 503 TGCAAC........................GTGAGAGAGGGAAAGGGGAGGG 531
     :::::: ||| :::||||||
 173 CysGlnGlyAspSerGlyGlyProValCysSerGlyLysLysLeuGlnGl 189

 532 C 532

                189 y 189

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-9

;   seq_documentation_block:
;   Sequence 9, Application US/0847167
;   Patent No. 5869302
;   GENERAL INFORMATION:
;   APPLICANT: LOOSMORE, Sheena M.
;   APPLICANT: YANG, Yan-Ping
;   APPLICANT: CHONG, Pele
;   APPLICANT: OOMEN, Raymond P.
;   APPLICANT: KLEIN, Michel H.
;   TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;   TITLE OF INVENTION: Reduced Protease Activity
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Sim & McBurney
;   STREET: Suite 701, 330 University Avenue
;   CITY: Toronto
;   STATE: Ontario
;   COUNTRY: Canada
;   ZIP: M5G 1R7
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/487,167
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/296,149
;   FILING DATE: 26-AUG-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/278,091
;   FILING DATE: 21-JUL-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Stewart, Michael I.
;   REGISTRATION NUMBER: 24,973
;   REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (416) 595-1155
;   TELEFAX: (416) 595-1163
;   INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 223 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-487-167-9

alignment_scores:
        Quality: 226.50   Length: 184
        Ratio: 2.137   Gaps: 6
Percent Similarity: 57.609   Percent Identity: 33.152

alignment_block:
US-09-030-606-174 x US-08-487-167-9

Align seg 1/1 to: US-08-487-167-9 from: 1 to: 223

   3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG 52
     |||||||||||||||||::::::::: :::||| :::
  37 SerAlaAlaHisCysTyrLysSerGlyIleGln..........ValAr 49

  53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
     ||||| ||| :::||:::|||:::   :
  49 gLeuGlyGluAspAsnIleAsnVal...ValGluGlyAsnGluGlnPheI 65

TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-146-2

alignment_scores:
Quality: 222.00   Length: 173
Ratio: 2.114   Gaps: 5
Percent Similarity: 60.694   Percent Identity: 31.792

alignment_block:
US-09-030-606-174 x US-08-557-146-2

Align seg 1/1 to: US-08-557-146-2 from: 1 to: 253

```
   3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCACACCATCGG  52
     ::!||||||||||||              :::::!||:||||:::||||
  67 ThrAlaAlaHisCys..........LysMetAsnGluTyrThrValHi  79

  53 GCTGGGCCTGCACAGTCTTGAGGCGCCACCAAGAGCCAAGAGCCAGATGG 102
     ||||||          ::!|||  :::!||:||||    :||||
  79 sLeuGlySerAspThrLeu...GlyAspArgArg.....AlaGlnArgI  93

 103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
     ::::!||||   ||| ||||||||||| |||::
  93 leLysAlaSerLysSerPheArgHisProGlyTyrSerThrGlnThrHis 109

 153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
     :!||||||||||||:!|||:||||||||||        |||:
 110 ValAsnAspLeuMetLeuIleLysLeuAsnSerGlnAlaArgLeuSerSe 126

 203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACT 252
     :                ::  :::!|||:!|||        ||:::::
 126 rMetValLysLysValArgLeuProSerArgCysGluProGlyThrT 143

 253 CTTGCTCCGTTCTTCGGCTGGGGTCTGCTGCGCGGCGTGAGCTCACGGGT 302
     :!||| ||||||||||||||||||
 143 hrCysThrValSerGlyTrpGly.........................  150

 303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCCAGTCGCGGGGGCTGAC 352
     ..........................................Th 151

 353 CCAGAGCTCTGCGGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
     ::!||     :::   :::!||:||     |||:::||||||:
 151 rThrThrSerProAspValThr..PheProSerAspLeuMetCysValAs 167

 403 CGTGTCGGTGGTGTCTGANGAGTCTGCANTAAGCTCTATGACCCGGCTGT 452
     :!||::::!||     ::: ||:::!||:||||||
 167 pValLysLeuIleSerProGlnAspCysThrLysValTyrLysAspLeuL 184

 453 ACCACCCCANCATGTTCTGCGCCGGCGAGGGCCAAGACCAGAAGGACTCC 502
     ::  :!||:||||||||     ::: :!||:||||||
 184 euGluAsnSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAla 200

 503 TGCAACGTGAGAGGGG 520
     ||||| ::!||
 201 CysAsnGlyAspSerGly 206
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND

```
 103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
     ::::!||||   ||| ||||||||||| |||::
  65 leSerAlaSerLysSerIleValHisProSerTyrAsnSerAsnThrLeu  81

 153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
     ||||||          ::!|||  :::!||:||||
  82 AsnAsnAspIleMetLeuIleLysLeuLysSerAlaAlaSerLeuAsnSe  98

 203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACT 252
     :   ::!|||||||::: ::::!||||| ::!||||||:::!|||:::
  98 rArgValAlaSerIleSerLeuProThrSerCysAlaSerAlaGlyThrG 115

 253 CTTGCCTCGTTCTGGCTGGGGTCTGCTGCGGCAACGGTGAGCTCACGGGT 302
     ::!||||:::!||||||||||||    |||  ::!||||
 115 lnCysLeuIleSerGlyTrpGly.......AsnThrLysSerSerGly 128

 303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCCAGTCGCGGGGGCTGAC 352
     .................................................
 129 ..ThrSer.......................................... 130

 353 CCAGAGCTCTGCGGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
     ||| |||||:::!||:::::
 131 .....................TyrProAspValLeuLysCysLeuLy 139

 403 CGTGTCGGTGGTGTCTGANGAGTCTGCANTAAGCTCTATGACCCGGCTGT 452
     ::: :::!||:||     |||   ::
 139 sAlaProIleLeuSerAspSerSerCysLysSerSerAlaTyrProGlyGlnI 156

 453 ACCACCCCANCATGTTCTGCGCCGGCGAGGGCCAAGACCAGAAGGACTCC 502
     ||||||||
 156 leThrSerAsnMetPheCysGlyAlaGlyTyrLeuGluGlyGlyLysAspSer 172

 503 TGCAAC...............GTGAGAGAGGGGAAAGGGGAGGG 531
     |||::              :::!||
 173 CysGlnGlyAspSerGlyGlyProValValCysSerGlyGlyLysLeuGlnGl 189

 532 C 532

 189 y 189
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2

seq_documentation_block:
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783

Wed Sep 29 14:27:20 1999

us-09-030-606-174.rai

```
;   TITLE OF INVENTION:  RELATED NUCLEIC ACIDS
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Eli Lilly and Company
;     STREET: Lilly Corporate Center
;     CITY: Indianapolis
;     STATE: Indiana
;     COUNTRY: United States of America
;     ZIP: 46285
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US96/04294
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/416,257
;     FILING DATE: 04-APR-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Blalock, Donna K.
;     REGISTRATION NUMBER: 38,082
;     REFERENCE/DOCKET NUMBER: X9239
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 317-277-1090
;     TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 253 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
PCT-US96-04294-2

alignment_scores:
        Quality: 222.00              Length:     173
        Ratio:   2.114               Gaps:         5
   Percent Similarity: 60.694    Percent Identity: 31.792

alignment_block:
US-09-030-606-174 x PCT-US96-04294-2

Align seq 1/1  to: PCT-US96-04294-2  from: 1  to: 253

  3 TCAGCGCCACACTGTTTCCAGGAAGTGAGTGCAGAGCTCCTACACCATCGG  52
    ::::||||||||||         :::::::||||||||::||||:::
 67 ThrAlaAlaHisCys..........LysMetAsnGluTyrThrValHi  79

 53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGG 102
    ||||||    ::||||||::   :::||::::::|    ::||||
 79 sLeuGlySerAspThrLeu...GlyAspArgArg.....AlaGlnArgI  93

103 TGGAGGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
    :::::||||||  ||||||||   ||||||  |||::
 93 leLysAlaSerLysSerPheArgHisProGlyTyrSerThrGlnThrHis 109

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCAGTCTGA 202
    :::|||||||||||||||::::||||||::::   |||:
110 ValAsnAspLeuMetLeuValLysLeuAsnSerGlnAlaArgLeuSerSe 126

203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACT 252
    :::::::::||||||   ::   |||:||| ||:::|||
126 rMetValLysValArgLeuProSerArgCysGluProGlyTyrThrT 143

253 CTTGCCTCCGTTTCTGGCTGGGGTCTGGGGTCGCCGAACGGTGAGCTCACGGGT 302
    ::|||  |||||||||||
143 hrCysThrValSerGlyTrpGly................ 150

303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGCGGGGCTGAC 352
```

```
151 ....................................................Th 151
                                                         ||
353 CCAGAGCTCGGTCCCAGGCAGAATGCCACCGTGCCACCGTGCTGCCAGTGCGTGAA 402
    |  :::||||     :::   ::|||    |||::||||||::
151 rThrThrSerProAspValThr..PheProSerAspLeuMetCysValAs 167

403 CGTGTCGTGGTGTCTGANGAGGTCTCANTAAGCTCTATGACCCGCTGT 452
    :||::|||||||       |||:::||||     ||||
167 pValLysLeuIleSerProGlnAspCysThrLysValTyrLysAspLeuL 184

453 ACCACCCCANCATGTTCTGCCGCCGGGGAGGGCAAGACCAGAAGGACTCC 502
    :::   :||||||||||||||    ::::||||:::|||:||
184 euGluAsnSerMetLeuCysAlaGlyIleProAspSerLysLysAsnAla 200

503 TGCAACGTGAGAGAGGGG 520
    ||||||     :::|||
201 CysAsnGlyAspSerGly 206
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-12

```
seq_documentation_block:
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
;   APPLICANT: Egelrud, Torbjorn
;   APPLICANT: Hansson, Lennart
;   TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
;   TITLE OF INVENTION: Enzyme (SCCE)
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: White & Case, Patent Department
;     STREET: 1155 Avenue of the Americas
;     CITY: New York
;     STATE: New York
;     COUNTRY: U.S.A.
;     ZIP: 10036-2787
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/557,146
;     FILING DATE: 14-DEC-1995
;     CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Sterner, Richard J.
;     REGISTRATION NUMBER: 35,372
;     REFERENCE/DOCKET NUMBER: 1103326-181
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 819-8783
;     TELEFAX:   (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 225 amino acids
;     TYPE: amino acids
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: polypeptide
US-08-557-146-12

alignment_scores:
        Quality: 218.00              Length:     173
        Ratio:   2.076               Gaps:         5
   Percent Similarity: 60.694    Percent Identity: 31.214

alignment_block:
US-09-030-606-174 x US-08-557-146-12

Align seq 1/1  to: US-08-557-146-12  from: 1  to: 225
```

REFERENCE/DOCKET NUMBER: 00398/100001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 286 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-467-155A-9

alignment_scores:
Quality: 210.50   Length: 232
Ratio: 1.949   Gaps: 8
Percent Similarity: 46.552   Percent Identity: 28.448

alignment_block:
US-09-030-606-174 x US-08-467-155A-9

Align seg 1/1 to: US-08-467-155A-9 from: 1 to: 286

```
   3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG  52
     ||||||||||||||||                           ::|||::
  69 SerAlaAlaHisCysTyrHisProGlnLeuGln........ValAr  81

  53 GCTGGGCCTGCACAGTCTTGAGGCC.................GACC  81
     ||||||      |||:|| ||::||                   ::
  81 gLeuGlyGluHisAsnArgAsnThrArgTyrProVal***IleTyrGluI  98

  82 AAGAGCCAGGAGCCAGATGGTGAGGCCGCTCCGTACGGCAACCCA  131
     ||       ||:|||||||:::|||||        :::  |||||
  98 leGluGlyAlaGluGlnPheIleAspAlaAlaLysMetIleLeuHisPro  114

 132 GAGTACAACAGACCCTTGCTCCGTAACGACCTCATCGTCATCAAGTTGGA  181
     ::|||::::     :::  ||||||:||||||||||||||
 115 AspTyrAspLysTrpThrValAspAsnAspIleMetLeuIleLysLeuLy  131

 182 CGAATCCGTGTCCGAGTCTGACACC................A  207
     |||||| |||                         :::||
 131 sSerProAlaThrLeuArgAsnThrArgTyrProVal***AsnSerLysV  148

 208 TCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAACTCTTGC  257
     ::::|||         |||:       ||||||||||||||:|::|||
 148 aIserThrIleProLeuProGlnTyrCysProThrAlaGlyThrGluCys  164

 258 CTCGTTCTTCGGTGGGGTCTGCTGGCGAACGGTGAGCTCACGGGTGTGTG  307
     |||||||||||||||:::|||    ||
 165 LeuValSerGlyTrpGlyValLeuLysPheGly...............  175

 308 TCTGCCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGTCGACCCAGA  357
     ||   ||||||||||||||||||||||||||||||||||||||||||||
 175 ...............................................  175

 358 GCTCTGCGTCCCAGGCAGAATG.CCTACCGTGCTGCAGTGCGTGAACCTG  406
                         ||| :|| :||:||||||||::||::::|
 176 ...........PheGluSerProSerValLeuGlnCysLeuAspAlaP  187

 407 TCGGTGGGTCT..................GANGAGGTCTGCAN  432
     |||:::|||                        ::|||||
 188 ProValLeuSerAspSerArgAsnThrArgTyrProVal***ValCysHi  204

 433 TAAGCTCTATGACCCGCTGTACCACCCCANCATGTTCTGCGCCGGCGGAG  482
                     |||||||||| ||    |||||||    |||
 204 sLysAlaTyrProArgGlnIleThrAsnAsnMetPheCysLeuGlyPheL  221

 483 GGCAAGACCAGAAGGACTCCTGCAACGTGAGAGAGGGAAAGGGGAGGGC  532
     ::|       |||||||||||::    :::||
 221 euGluGlyLyLysAspSerCysGlnTyrAspSerGly...........  233
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-9

seq_documentation_block:
; Sequence 9, Application US/08467155A
; Patent No. 5736377
; GENERAL INFORMATION:
; APPLICANT: Band, Vimla
; TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
; TITLE OF INVENTION: MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/467,155A
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Clark, Paul T.
;   REGISTRATION NUMBER: 30,162

```
533 AGGCGACTCAGGGAGGAGGGTGGAGAAGGGGGAGACAGAGACACAGGGCC 582
                                                   |||||
234 ...............................................GlyPr 235

583 G........CATGGCGAGATGCAG........AGATGGAGAGAC 610
    |    :::|||||:::|||          |||  |||
235 oValValCysAsnGlyGluValGlnGlyIleValSerTrpGlyAsp 250
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-628-198-9

seq_documentation_block:
; Sequence 9, Application US/08628198
; Patent No. 5843694
; GENERAL INFORMATION:
;   APPLICANT: Band, Vimla
;   TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
;   TITLE OF INVENTION: MOLECULES AND METHODS
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/628,198
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/467,155
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 514
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clark, Paul T.
;   REGISTRATION NUMBER: 30,162
;   REFERENCE/DOCKET NUMBER: 00398/100002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
;   TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 286 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-628-198-9

alignment_scores:
        Quality: 210.50   Length: 232
        Ratio: 1.949   Gaps: 8
Percent Similarity: 46.552   Percent Identity: 28.448

alignment_block:
US-09-030-606-174 x US-08-628-198-9

Align seg 1/1 to: US-08-628-198-9 from: 1 to: 286

```
  3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG 52
    ||||||||||||||:::::||:::||              :::
 69 SerAlaAlaHisCysTyrHisProGlnLeuGln............ValAr 81

 53 GCTGGGCCTGCACAGTCTTGAGGCC....................GACC 81
    |||||||||:::||:  :::||                        :::
 81 gLeuGlyGluHisAsnArgAsnThrArgTyrProVal***IleTyrGluI 98
```

```
 82 AAGAGCCAGGAGCCAGCAGATGTGGAGGCCAGCCTCCGTACGGCCACCCA 131
    |||                      :::||||:::   :::   |||||
 98 leGluGlyAlaGluGlnPheIleAspAlaAlaLysMetIleLeuHisPro 114

132 GAGTACAACAGACCCTTGCTGCTAACGACCTCATCGTCATCAAGTTGGA 181
    :::||::::::|:::      :::   |||||:::||||||||||||||||
115 AspTyrAspLysTrpPheValAspAsnAspIleMetLeuIleLysLeuLy 131

182 CGAATCCGTGTCCGAGTCTGACACC.................A 207
    :::||:::||||||          :::||
131 sSerProAlaThrLeuArgAsnThrArgTyrProVal***AsnSerLysV 148

208 TCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGGAACTCTTGC 257
    :::   :::||     |||||||:::||||:::|||||||||:::||
148 alSerThrIleProLeuProGlnTyrCysProThrAlaGlyThrGluCys 164

258 CTCGTTTCTGCTGGGGTCTCTGGCAACGTGAGCTCACGGGTGTGTG 307
    |||||||||||||||:::||:::|||
165 LeuValSerGlyTrpGlyValLeuLysPheGly.............. 175

308 TCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGACCCAGA 357
    ................................................ 175

358 GCTCTGCGTCCCAGGCAGAATG.CCTACCGTCGCTGCCAGTCGTGAACGTG 406
    :::||:::||||||||||||:::|||||||||||:::|||||||||::|
176 ...........PheGluSerProSerValLeuGlnCysLeuAspAla 187

407 TCGGTGCGTGTCT...........................GANGAGGTCGCAN 432
    |||:::|||                              ::|||||
188 ProValLeuSerAspSerArgAsnThrArgTyrProVal***ValCysHi 204

433 TAAGCTCTATGACCCGCTGACCACCCCANCATGTTCTGCGCCGCGGCGAG 482
    ||||||||        |||        :::||
204 sLysAlaTyrProArgGlnIleThrAsnAsnMetPheCysLeuGlyPheL 221

483 GGCAAGACCAGAAGGACTCCTGCAACGTGAGGAGGGGAAAGGGGAGGGGC 532
    :::        ||||||||||||||||   :::||
221 euGluGlyGlyLysAspSerCysGlnTyrAspSerGly.......... 233

533 AGGCGACTCAGGGAGGAGGGTGGAGAAGGGGGAGACAGAGACACAGGGCC 582
                                                   |||||
234 ...............................................GlyPr 235

583 G........CATGGCGAGATGCAG........AGATGGAGAGAC 610
    |    :::|||||:::|||          |||  |||
235 oValValCysAsnGlyGluValGlnGlyIleValSerTrpGlyAsp 250
```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US96-07343-9

seq_documentation_block:
; Sequence 9, Application PC/TUS9607343
; GENERAL INFORMATION:
;   APPLICANT: New England Medical Center Hospitals, Inc.
;   TITLE OF INVENTION: NES-1 POLYPEPTIDES, DNA, AND RELATED
;   TITLE OF INVENTION: MOLECULES AND METHODS
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US96/07343
;   FILING DATE:
;   CLASSIFICATION:

```
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/467,155
    FILING DATE: 06-JUN-1995
    CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
    NAME: Clark, Paul T.
    REGISTRATION NUMBER: 30,162
    REFERENCE/DOCKET NUMBER: 00398/100001
TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617/542-5070
    TELEFAX: 617/542-8906
    TELEX: 200154
INFORMATION FOR SEQ ID NO: 9:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 286 amino acids
    TYPE: amino acid
    STRANDEDNESS: not relevant
    TOPOLOGY: linear
    MOLECULE TYPE: protein
PCT-US96-07343-9
```

```
alignment_scores:
        Quality: 210.50                Length:    232
          Ratio:   1.949                 Gaps:      8
Percent Similarity: 46.552       Percent Identity: 28.448

alignment_block:
US-09-030-606-174 x PCT-US96-07343-9     ..

Align seg 1/1  to: PCT-US96-07343-9  from: 1  to: 286
```

```
   3 TCAGCCGCACACTGTTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG  52
     ||||||||||||||||||||::::||            ::||
  69 SerAlaAlaHisCysTyrHisProGlnLeuGln..........ValAr  81

  53 GCTGGGGCTGCACAGTCTTGAGGCC....................GACC  81
     ||||||  |||::||                               :::|
  81 gLeuGlyGluHisAsnArgAsnThrArgTyrProVal***IleTyrGluI  98

  82 AAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCA 131
     |||  :::||||||||||||:::::||:::   ::|   |||||
  98 leGluGlyAlaGluGlnPheIleAspAlaAlaLysMetIleLeuHisPro 114

 132 GAGTACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGA 181
     ::||:::||:::  :::  |||||::|||||||||||||||
 115 AspTyrAspLysTrpThrValAspAsnAspIleMetLeuIleLysLeuLeu 131

 182 CGAATCCCGTGTCCGAGTCTGACACC..................A 207
     ::||||              ::|||
 131 sSerProAlaThrLeuArgAsnThrArgTyrProVal***AsnSerLysV 148

 208 TCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAACTCCTTGC 257
     ::  :::||        :::  |||||||||||||:::::|||
 148 aIleSerThrIleProLeuProGlnTyrCysProThrAlaGlyThrGluCys 164

 258 CTCGTTCGGCGTCGGGGTCTGCTGGCGAACGGTGAGCTCACGGGTGTGTG 307
     ||||||||||||||||:||||       |||
 165 LeuValSerGlyTrpGlyValLeuLysPheGly.............. 175

 308 TCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGACCCAGA 357
 175 .............................................. 175

 358 GCTCTGCGGTCCCAGGCAGAATG.CCTACCGTGCTGCAGTGCCTGAACGTG 406
     ||| :||:::||||||||||||::::::::::
 176 ...........PheGluSerProSerValLeuGlnIleCysLeuAspAla 187

 407 TCGGTGGTGTCT...................GANGAGGTCTGCAN 432
     ||::::|||              :::|||||||
 188 ProValLeuSerAspSerArgAsnThrArgTyrProVal***ValCysHi 204

 433 TAAGCTCTATGACCCGGCTGTACCACCCCANCATGTTCTCGCGCCGGCGGAG 482
```

Second column:

```
   || ||| ||                         |||||||||| |||
 204 sLysAlaTyrProArgGlnIleThrAsnAsnMetPheCysLeuGlyPheL 221

 483 GGGAAGACCAGAGGACTCCTGCAACGTGAGAGAGGGGAAAGGGGAGGGC 532
                      |||||||||||||||: ::||
 221 euGluGlyGlyLysAspSerCysGlnTyrAspSerGly............ 233

 533 AGGCGACTCAGGGAGGGTGGAGAAGGGGAGCAGAGACACAGCACAGGGCC 582
                                             ||||||
 234 ..............................................GlyPr 235

 583 G........CATGGCGAGATGCAG........AGATGGAGAGAC 610
           ::||||||:::||:::||       ||| |||
 235 oValValCysAsnGlyGluValGlnGlyIleValSerTrpGlyAsp 250
```

```
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-10

seq_documentation_block:
; Sequence 10, Application US/08096946
; Patent No. 5516639
; GENERAL INFORMATION:
;    APPLICANT: Tindall, Donald J.
;    APPLICANT: Young, Charles Y-F
;    APPLICANT: Klee, George G.
;    TITLE OF INVENTION: Antibodies Specific for Human
;    TITLE OF INVENTION: Prostate Glandular Kallikrein
;    NUMBER OF SEQUENCES: 11
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Merchant & Gould
;       STREET: 3100 No. 5516639west Center
;       CITY: Minneapolis
;       STATE: MN
;       COUNTRY: USA
;       ZIP: 55402
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/096,946
;       FILING DATE: 19930722
;       CLASSIFICATION: 436
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Woessner, Warren D.
;       REGISTRATION NUMBER: 30,440
;       REFERENCE/DOCKET NUMBER: 1016.62-US-01
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 612-332-5300
;       TELEFAX: 612-332-8081
;    INFORMATION FOR SEQ ID NO:  10:
;       SEQUENCE CHARACTERISTICS:
;       LENGTH: 237 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       IMMEDIATE SOURCE:
;       CLONE: Amino acid sequence of hK2
;       CLONE: (deduced from cDNA sequence)
US-08-096-946-10
```

```
alignment_scores:
        Quality: 191.50                Length:    184
          Ratio:   1.974                 Gaps:      6
Percent Similarity: 52.717       Percent Identity: 29.891

alignment_block:
US-09-030-606-174 x US-08-096-946-10     ..

Align seg 1/1  to: US-08-096-946-10  from: 1  to: 237
```

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Raasch, Kevin W.
REGISTRATION NUMBER: 35,651
REFERENCE/DOCKET NUMBER: 150.62WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: Amino acid sequence of hK2
CLONE: (deduced from cDNA sequence)
PCT-US94-07329-10

alignment_scores:
Quality: 191.50   Length: 184
Ratio: 1.974   Gaps: 6
Percent Similarity: 52.717   Percent Identity: 29.891

alignment_block:
US-09-030-606-174 x PCT-US94-07329-10

Align seg 1/1   to: PCT-US94-07329-10   from: 1 to: 237

```
  3  TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACCACCATCGG  52
     :::||||||||||||||||::::|||                ::::
 38  ThrAlaAlaHisCysLeuLysLysAsnSerGln.........ValTr     50

 53  GCTGGGCCTGCACAGTCTTGAGGCCCACCAAGAGCCAGGAGCCAGGATGG 102
     ||||||   |||:::||     :::::  ||| ||| |
 50  pLeuGlyArgHisAsnLeuPheGluProGluAspThrGly...GlnArgV  66

103  TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
     ||  :::||| ||| :|||||  ||||||  |||||| ||||||||
 66  alProValSerHisSerPheProHisProLeuTyrAsnMetSerLeuLeu  82

153  ........................................GCTAACGACCTCATGCT 169
                                              :::::|||||||||||||
 83  LysHisGlnSerLeuArgProAspGluAspSerSerHisAspLeuMetLe  99

170  CATCAAGTTGGACGAATCCGTGTCCAGTCTGACACCATCCGGAGCATCA 219
     |::::||::||||| :::: :|||:::|||::||||::::::    :::
 99  uLeuArgLeuSerGluProAlaLysIleThrAspValValLysValLeuG 116

220  GCATTGCTTCGCAGTGCCCTACCGCGGGAACTCTTGCCTCGTTTCTGGC 269
     ||::: :::|||:: ::|||||||  |||:::|::|:::|||::||||||
116  lyLeuProThrGlnGluProAlaLeuGlyThrThrCysTyrAlaSerGly 132

270  TGGGGTCTGCTGGCAAACGGTGAGCTCACGGGTGTGTCTCGCCCTCTTC 319
     ||||||    :::  |||
133  TrpGlySerIle..Glu.................................. 137

320  AAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGACCAGGAGCTCTGCCGTCCC 369
     :::|||       |||:||||||||
138  .........................ProGluPheLeu.. 142

370  AGGCAGAATGCCTACCGTGCTGCAGTGCGTGAACGTGCGTGGTGTCTG 419
     |:::::||||||||:::::|||||||||||||::||:: :::||
143  ......ArgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerA 157

420  ANGAGGTCTGCANTAAGCTCTATGACCCGCTGTACCACCCANCATGTTC 469
     :::||:::||:::|| |||:::||||| |||:::
157  snAspMetCysAlaArgAlaTyrSerGluLysValThrGluPheMetLeu 173

470  TGCGCCGGCGGAGGGCAAGACCAGAAGGACTCCTGCAACGTGAGAGAGGG 519
     ||::::||||    :::||
174  CysAlaGlyLeuTrpThrGlyGlyLysAspThrCysGlyGlyAspSerGl 190

520  G 520

190  y 190
```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-07329-10

seq_documentation_block:
; Sequence 10. Application PC/TUS9407329
; GENERAL INFORMATION:
;   APPLICANT: Mayo Foundation for Medical
;   APPLICANT: Education and Research
;   TITLE OF INVENTION: Antibodies Specific for Human
;   TITLE OF INVENTION: Prostate Glandular Kallikrein
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Schwegman, Lundberg & Woessner
;   STREET: 3500 IDS Center
;   STREET: 80 South Eighth Street
;   CITY: Minneapolis
;   STATE: MN
;   COUNTRY: USA
;   ZIP: 55402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US94/07329

Wed Sep 29 14:27:20 1999

us-09-030-606-174.rai

```
        |||||||||    ||||||::::||:::||    ::::||
174 CysAlaGlyLeuTrpThrGlyGlyLysAspThrCysGlyGlyAspSerGl 190

520 G 520
    |
190 y 190
```

```
              GenCore version 4.5
       Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     September 25, 1999, 07:29:47 ; Search time 165.01 Seconds
                                           (without alignments)
                                  648.364 Million cell updates/sec

Title:          US-09-030-606-175
Perfect score:  1167
Sequence:       1 GGCGAGCCCTGGCAGGCGGC.........NTAAGAGAAGCGCAAAAAA 1167

Scoring table:  IDENTITY_NUC

Searched:       176461 seqs, 45838279 residues

Database :      Issued_Patents_NA:*
                1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
                4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
                5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
                6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	721.4	61.8	871	3	US-08-744-026-2	Sequence 2, Appli
2	181.6	15.6	8174	1	US-07-914-281-5	Sequence 5, Appli
3	181.6	15.6	8174	1	US-08-393-246-5	Sequence 5, Appli
4	181.6	15.6	8174	5	US-08-525-058A-5	Sequence 5, Appli
5	181.6	15.6	8174	5	PCT-US91-00899-3	Sequence 3, Appli
6	154.6	13.2	1089	5	PCT-US96-04294-1	Sequence 1, Appli
7	154.6	13.2	1089	5	PCT-US96-04294-3	Sequence 3, Appli
8	153	13.1	986	4	US-08-557-146-1	Sequence 1, Appli
9	142.2	12.2	4060	1	US-08-308-949A-1	Sequence 1, Appli
10	140	12.0	732	3	US-08-361-395-2	Sequence 2, Appli
11	125.4	10.7	833	4	US-08-790-137-2	Sequence 2, Appli
12	95.8	8.2	957	3	US-08-684-862-11	Sequence 11, Appl
13	85.8	7.4	1462	2	US-08-358-782D-14	Sequence 14, Appl
14	84.2	7.2	992	2	US-08-358-782D-13	Sequence 13, Appl
15	84.2	7.2	1729	4	US-08-844-024-1	Sequence 1, Appli
16	81	6.9	832	5	PCT-US95-06157-5	Sequence 5, Appli
17	81	6.9	760	5	PCT-US95-06157-7	Sequence 7, Appli
18	81	6.9	766	5	PCT-US95-06157-9	Sequence 9, Appli
19	80	6.9	1454	4	US-08-467-155A-2	Sequence 2, Appli
20	80	6.9	1454	4	US-08-628-198-2	Sequence 2, Appli
21	80	6.9	1454	5	PCT-US96-07343-2	Sequence 2, Appli
22	75.4	6.5	699	3	US-08-738-413B-8	Sequence 8, Appli
23	71.2	6.1	840	3	US-08-684-862-12	Sequence 12, Appl
24	70.4	6.0	1333	3	US-08-684-862-9	Sequence 9, Appli
25	66	5.7	988	3	US-08-684-862-10	Sequence 10, Appl
26	65.6	5.6	1096	3	US-08-684-862-8	Sequence 8, Appli
27	62.4	5.3	7218	2	US-08-232-463-14	Sequence 14, Appl
28	57.8	5.0	734	3	US-08-650-129-1	Sequence 1, Appli
29	57.8	5.0	821	3	US-08-650-129-2	Sequence 2, Appli
30	57.8	5.0	866	3	US-08-650-129-3	Sequence 3, Appli
31	54	4.6	2259	4	US-08-845-998-3	Sequence 3, Appli
32	52.4	4.5	2218	4	US-08-845-998-5	Sequence 5, Appli
33	51.4	4.4	200	3	US-08-931-981A-1	Sequence 1, Appli
34	49	4.2	1554	5	US-08-469-486-1	Sequence 1, Appli
35	47.2	4.0	8174	1	US-07-914-281-5	Sequence 5, Appli
36	47.2	4.0	8174	1	US-08-393-246-5	Sequence 5, Appli
37	47.2	4.0	8174	3	US-08-525-058A-5	Sequence 5, Appli
38	47.2	4.0	8174	5	PCT-US91-00899-3	Sequence 3, Appli
39	46.8	4.0	884	4	US-08-851-974-2	Sequence 2, Appli
40	46	3.9	329	1	US-08-148-910-2	Sequence 2, Appli
41	46	3.9	970	1	US-08-148-910-3	Sequence 3, Appli
42	46	3.9	329	1	US-08-148-910-13	Sequence 13, Appl
43	46	3.9	2033	1	US-08-148-910-14	Sequence 14, Appl
44	46	3.9	329	2	US-08-448-937A-2	Sequence 2, Appl
45	46	3.9	970	2	US-08-448-937A-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-08-744-026-2

; Sequence 2, Application US/0874026
; Patent No. 5780148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-08-744-026-2

Query Match              61.8%; Score 721.4; DB 3; Length 871;
Best Local Similarity    98.2%; Pred. No. 1.2e-184;
Matches 725; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy    2 CGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTCGGGGCGTCCTGG 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134 CCCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTCGGGGCGTCCTGG 193

Qy   62 TGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTTCCAGAACTCCTACACCATCGGGC 121
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194 TGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTTCCAGAACTCCTACACCATCGGGC 253
```

ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-07-914-281-5

Query Match   15.6%;   Score 181.6; DB 1;   Length 8174;
Best Local Similarity 77.3%;   Pred No. 1.2e-39;
Matches 269; Conservative   0; Mismatches 67; Indels 12; Gaps 4;

```
Qy   700 ATTCAGGAATATCGTTCCCAGGCCCCTCCCTCCCTCAGGCCCAGGAGTCCAGGCCCCCAGC 759
         | | ||||  |||| | | ||| ||||||| || |||||||| |||||||| |||| ||||
Db  1025 ACTTAGGAGTCCAGGTCCCGCGGCCCCTCCCTTCCTCAGACCCAGAGTCCAAGCCCCCTGC 966

Qy   760 CCCTCCTCCCTCAAACCAAGG-GTACAGATCCCAGCCCTGCTCCTCCCTCAGACCCAGGAG 818
         |||||| ||| | |   ||||  |||  ||||||||||||||| ||| ||||||| |||||
Db   965 CCCTCCTTCCTCAGACCCCAGGACCCCAGCCCCTGCTCCTCCTTCCTCACACCCACGAG 906

Qy   819 TCCAGACCCCCCAGCCCTCCNTCCNTCAGCACCCAGGAGTCCAG------CCCTCCTC 870
         ||||||    |  |||||| ||| ||||||||||||||||||||       |||||||
Db   905 TCCAGA-TCCCTAGCCCTCCTACTCCCTCAGCACCCAGGAGTCCAGACCAAAGCTCCCTCCTC 847

Qy   871 CNTCAGACGCAGGAGTCCAGACCCCCCAGCCCNTCCGTCAGACCCCGGGTGCAGGC 930
         | ||||| ||| | |||| ||| ||||||| || ||| ||||||||||||| ||||
Db   846 CCTCAGACCCAGGAGCCCAAGTCCCAGAGTCCCCAGCCCCCTCCTCCCTCAGATCCAGGAGTACAGG- 787

Qy   931 CCCCAACCCTCNTCNTCAGAGTCAGGAGGTCCAAGCCCCTCGTTCCCCAGAC 990
         |||| |||| ||||| ||| |  |  |  || ||| ||||||| |||| ||||
Db   787 -CCCAGACCCTCCTCCCTCAGACCCAGGAGTCCAGGCCCCCACCCCTCCTCCCTCCCTCAGAC 729

Qy   991 CCAGAGTGNCAGGTCCAGGCCCCTCCCTCCCTCAGACCCAGCCGGTCCAA 1038
         |||| | || ||| ||||||||||||||||| ||| ||||||||||||||
Db   728 CCAGGAGTCCAGACCCCAGGCCCCTCCCTCCCTCAGACCAGCAGAAGGCCTA 681
```

RESULT 3
US-08-393-246-5/c
; Sequence 5, Application US/08393246
; Patent No. 5595900
; GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```
Qy   122 TGGGCCTGCACAGTCTTGAGGCGACCAAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCC 181
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254 TGGGCCTGCACAGTCTTGAGGCGACCAAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCC 313

Qy   182 TCTCCGTACGGCACCCAGATGACAACAGACTCTTGCTCGCTCAACGACCTCATGCTCATCA 241
         ||||||||||||||||||||||||||||||||||||||| ||||| ||||||| |||||
Db   314 TCTCCGTACGGCACCCAGATGACAACAGACTCTTGCTCGCTAACGACCTCATGNTCATCA 373

Qy   242 AGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGT 301
         ||||||| ||||||||||||||||||||| |||||||||| |||  ||  |||| ||||
Db   374 AGTTGGAGGAATCCGTGTCCGAGTCTGACAACATCCGGAGNATCCGGAACGGCAGAA 433

Qy   302 GCCCTACCGCGGGGAACTCTTGCCTCGTNTCGGCTGGTCTGCTCGTGGCGAACGGCAGAA 361
         |||||||| |||||| ||||||||   |  |||||||||||||| |||||||||||||||
Db   434 GCCCTACCGCGGGGAACTTTGCCTCGTTCTGGCTGGGTCTGCTGCTCGTGGCGAACGGCAGAA 493

Qy   362 TGCCTACCGTGCTGCACTCGCGTGAACGTGTCGGTGGTGTCTGGAGGANGTCTGCAGTAAGC 421
         ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   494 TGCCTACCGTGCTGCAGTCGCGTGAACGTGTCGGTGGTGTCTGGAGGGTCTGCAGTAAGC 553

Qy   422 TCTATGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGTCGGGGCCAAGACCAGAAGG 481
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   554 TCTATGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGTCGGGGCCAAGACCAGAAGG 613

Qy   482 ACTCCTGCAACGGTCGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGGCCTTG 541
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   614 ACTCCTGCAACGGTCGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGGCCTTG 673

Qy   542 TGTCTTTCGGAAAAAGCCCGTGTGGCCAACTTGGCGTGCCAGGTGTCTACACCAACCTCT 601
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   674 TGTCTTTCGGAAAAGCCCGTGTGGCCAAGTTGGCGTGCCAGGTGTCTACACCAACCTCT 733

Qy   602 GCAAATTCACTGAGTGGATGAGATAGAGAAAACCGTCCAGNCCAGTTAACTCTGGGGACTGGGAA 661
         |||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db   734 GCAAATTCACTGAGTGGATGAGATAGAGAAAACCGTTCCAGGCCAGTTAACTCTGGGGACTGGGAA 793

Qy   662 CCCATGAAATTGACCCCCAAATACATCCTGCGGAANGAATTCAGGAATATCTGTTCCCAG 721
         ||||||||||||||||||||||||||||||| |  ||||||||||||||||||||||||
Db   794 CCCATGAAATTGACCCCCAAATACATCCTGCGGAAGGAATTCAGGAATATCTGTTCCCAG 853

Qy   722 CCCTCCTCCCTCAGGCC 739
         |||||||||||||||||
Db   854 CCCTCCTCCCTCAGGCC 871
```

RESULT 2
US-07-914-281-5/c
; Sequence 5, Application US/07914281
; Patent No. 5324663
; GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/914,281
FILING DATE: 19920720
CLASSIFICATION: 530

```
;   APPLICATION NUMBER: US/08/393,246
;   FILING DATE:
;   CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/220,433
;   FILING DATE: 30-MAR-1994
;   APPLICATION NUMBER: US 07/914,281
;   FILING DATE: 20-JUL-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: Lavalleye, Jean-Paul M. P.
;   REGISTRATION NUMBER: 31,451
;   REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (703)521-4500
;   TELEFAX:  (703)486-2347
;   TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  8174 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS: unknown
;     TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
;   ANTI-SENSE: NO
; US-08-393-246-5

Query Match             15.6%;     Score 181.6;  DB 1;  Length 8174;
Best Local Similarity   77.3%;     Pred. No. 1.2e-39;
Matches  269; Conservative   0; Mismatches  67; Indels  12; Gaps   4;

Qy   700 ATTCAGGAAATATCTGTTCCAGCCCCTCCTCCCTCAGGCCCCAGGAGTCCAGGCCCCCAGC 759
           |  | ||||  | |||| ||||||||||||||||||| ||| |||||||||||||| |||| ||
Db  1025 ACTTAGGAGTCCAGGCTCCCGGCCCCCGGCCCCTCCTCAGACCCAGGAGTCCAAGCCCCCTGC 966

Qy   760 CCCTCCTCCCTCAAACCAAGG-GTACAGATCCCAGACCCCTCCTCCCTCAGACCCAGGAG 818
           ||||||| ||| | |||  ||  ||| |  |||||| ||||||||||| ||||| |||
Db   965 CCCTCCTTCCTCAGACCCAGGAGTCCAGGACCCCAGCCCCCCTTCCTCACACCCACGAG 906

Qy   819 TCCAGACCCCCCAGCCCCTCNTCCNTCAGACCAGGAGTCCAG------CCCTCCTC 870
           ||||| ||||| ||||||||| || |||||||||||||||||      |||||||||
Db   905 TCCAGA-TCCCTAGCCCCTACTCCCTCAGACCCTCAGACCAGGAGTCCAGACCAAAGCTCCCTCCTC 847

Qy   871 CNTCAGACGCAGGAGTCCAGACCCCCNTCNTCCGTCAGACCCAGGGGTGCAGGC 930
           | |||||| ||||||||||| ||||| || | ||||||||||||||| ||||
Db   846 CCTCAGACCCAGGAGCCCAAGTCCCCAGCCCCTCCCCAGATCCAGATTCAGGAGTACAGG- 787

Qy   931 CCCCAACCCCTCNTCNTCAGAGTCAGAGGTCCAGGCCCCAACCCCTCGTTCCCCAGAC 990
           |||| ||||| || | ||||| | ||| ||||||||||| ||||| | |  ||||
Db   787 -CCCAGACCCTCCTCCCTCAGACCCAGGAGTCCAGGCCCCCCACCCCTCCTCCCTCAGAC 729

Qy   991 CCAGAGGTNCAGGTCCAGCCCCTCCCAGCCCCTCCCAGCCCCTCCTCCCTCAGACCCAGGCGGTCCAA 1038
           ||||| ||  ||| |||||||||||||| ||||||||||| |||||||| || | ||||
Db   728 CCAGGAGTCCAGAGCCCCAGCCCCTCCTCCCTCCCTCAGACACAGAAGGCCTA 681
```

RESULT 4
US-08-525-058A-5/c

; Sequence 5, Application US/08525058A
; Patent No. 5770420
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
;   TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
;   TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
;   TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
;   TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
;   STREET: 1755 Jefferson Davis Highway, Fourth Floor
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.

;   ZIP: 22202
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/525,058A
;   FILING DATE:
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Lavalleye, Jean-Paul M. P.
;   REGISTRATION NUMBER: 31,451
;   REFERENCE/DOCKET NUMBER: 2363-060-55
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)521-4500
;   TELEFAX: (703)486-2347
;   TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 8174 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   ANTI-SENSE: NO
; US-08-525-058A-5

Query Match   15.6%;   Score 181.6; DB 3; Length 8174;
Best Local Similarity   77.3%;   Pred. No. 1.2e-39;
Matches 269; Conservative   0; Mismatches 67; Indels 12; Gaps   4;

Qy   700 ATTCAGGAAATATCTGTTCCAGCCCCTCCTCCCTCAGGCCCCAGGAGTCCAGGCCCCCAGC 759
           | | |||| | |||| ||||||||||||||||||| ||| |||||||||||||| |||| ||
Db 1025 ACTTAGGAGTCCAGGCTCCCGGCCCCCGGCCCCTCCTCAGACCCAGGAGTCCAAGCCCCCTGC 966

Qy   760 CCCTCCTCCCTCAAACCAAGG-GTACAGATCCCAGACCCCTCCTCCCTCAGACCCAGGAG 818
           ||||||| ||| | ||| || ||| | |||||| ||||||||||| ||||| |||
Db   965 CCCTCCTTCCTCAGACCCAGGAGTCCAGGACCCCAGCCCCCCTTCCTCACACCCACGAG 906

Qy   819 TCCAGACCCCCCAGCCCCTCNTCCNTCAGACCAGGAGTCCAG------CCCTCCTC 870
           ||||| ||||| ||||||||| || ||||||||||||||||| |||||||||
Db   905 TCCAGA-TCCCTAGCCCCTACTCCCTCAGACCCTCAGACCAGGAGTCCAGACCAAAGCTCCCTCCTC 847

Qy   871 CNTCAGACGCAGGAGTCCAGACCCCCNTCNTCCGTCAGACCCAGGGGTGCAGGC 930
           | |||||| ||||||||||| ||||| || | ||||||||||||||| ||||
Db   846 CCTCAGACCCAGGAGCCCAAGTCCCCAGCCCCTCCCCAGATCCAGATTCAGGAGTACAGG- 787

Qy   931 CCCCAACCCCTCNTCNTCAGAGTCAGAGGTCCAGGCCCCAACCCCTCGTTCCCCAGAC 990
           |||| ||||| || | ||||| | ||| ||||||||||| ||||| | | ||||
Db   787 -CCCAGACCCTCCTCCCTCAGACCCAGGAGTCCAGGCCCCCCACCCCTCCTCCCTCAGAC 729

Qy   991 CCAGAGGTNCAGGTCCAGCCCCTCCCAGCCCCTCCCAGCCCCTCCTCCCTCAGACCCAGGGGTCCAA 1038
           ||||| || ||| |||||||||||||| ||||||||||| |||||||| || | ||||
Db   728 CCAGGAGTCCAGAGCCCCAGCCCCTCCTCCCTCCCTCAGACACAGAAGGCCTA 681
```

RESULT 5
PCT-US91-00899-3/c

; Sequence 3, Application PC/TUS9100899
; GENERAL INFORMATION:
; APPLICANT: Lowe, John B.
;   TITLE OF INVENTION: Method and Products For the Synthesis of
;   TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,
;   TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned
;   TITLE OF INVENTION: Genetic Sequences That Determine These Structur
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;   ADDRESSEE: P.C.
;   STREET: 1755 Jefferson Davis Highway, Suite 400
;   CITY: Arlington

```
        STATE:  Virginia
        ZIP:  22202
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER:  IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: PCT/US91/00899
        FILING DATE: 19910214
        CLASSIFICATION:  435
    ATTORNEY/AGENT INFORMATION:
        NAME:  Lavalleye Ph.D., Jean-Paul
        REGISTRATION NUMBER:  31,451
        REFERENCE/DOCKET NUMBER:  2363-021-55 PCT
    TELECOMMUNICATION INFORMATION:
        TELEPHONE:  (703)521-5940
        TELEFAX:  (703)486-2347
        TELEX:  248855 OPAT UR
    INFORMATION FOR SEQ ID NO:  3:
    SEQUENCE CHARACTERISTICS:
        LENGTH:  8174 base pairs
        TYPE:  NUCLEIC ACID
        STRANDEDNESS:  unknown
        TOPOLOGY:  unknown
    MOLECULE TYPE:  cDNA
    HYPOTHETICAL:  NO
    ORIGINAL SOURCE:
        ORGANISM:  Homo sapiens
        TISSUE TYPE:  Blood
    FEATURE:
        NAME/KEY:  misc_feature
        LOCATION:  4686..5780
        OTHER INFORMATION:  /label= mat_peptide
PCT-US91-00899-3
```

```
Query Match        15.6%;  Score 181.6;  DB 5;  Length 8174;
Best Local Similarity  77.3%;  Pred. No. 1.2e-39;
Matches 269;  Conservative    0;  Mismatches   67;  Indels   12;  Gaps   4;

Qy   700 ATTCAGGAATATCTGTTCCCAGCCTCCTCCCTCAGGCCCAGGAGTCCAGGCCCCCAGC 759
             | ||||| | ||||| ||||||||||| |||| |||||||||| ||||||||||||||
Db  1025 ACTTAGGAGTCCAGGTCCCGGCCCCTCCTTCTCAGACCCAGGAGTCCAAGCCCCCTGC 966

Qy   760 CCCTCCTCCCTCAAACCAAGG-GTACAGATCCCAGACCCCTCCTCCCTCAGACCCAGGAG 818
             |||||||||||| || ||||| | |||| |||||||||||||| ||||||||||||| |
Db   965 CCCTCCTTCCTCAGACCCAGGAGTCCAGGACCCCAGACCCCTCCTTCCTCCACCCACGAG 906

Qy   819 TCCAGACCCCCCAGCCCTCNTCCNTCAGACCCAGGAGTCCAG-------CCCTCCTC 870
             ||||| |||||||||||| | |||||||||||||||            |||||||||
Db   905 TCCAGA-TCCCTAGCCCCCTACTCCCTCAGACCCAGGAGTCCAGACCAAAGCTCCCTCCTC 847

Qy   871 CNTCAGACGCAGGAGTCCAGACCCCAGCCCNTCNTCCGTCAGACCCAGGGGTCCAGGC 930
             | ||||||| ||||||||||| ||||||| |  |||||| ||||||||||| |||| |
Db   846 CCTCAGCAGCCAGGAGCCCAAGTTCCCAGCCCCTCCCAGCCCTCCCAGATCCAGGAGTACAGG- 787

Qy   931 CCCCAACCCCTCNTCNTCAGAGTCAGAGGTCCAAGCCCCCAACCCTCGTTCCCCAGAC 990
             ||| ||||||| || |||| ||||||| || ||||||||||||||| || |||||||
Db   787 -CCCGACCCCTCCTCAGACCCAGGAGTCCAGGCCCCCACCCCTCCTCCCTCCCTCAGAC 729

Qy   991 CCAGAGGTNCAGGTCCCAGCCCCTCCTCCCTCAGACCCCAGCGGTCCAA 1038
             ||||||| |||| |||||||||||||| ||||||||| ||| | ||||
Db   728 CCAGGAGTCCAGAGCCCCAGCCCCTCCTCCCTCCCCAGACACAGAGGCCTA 681
```

```
RESULT   6
PCT-US96-04294-1/c
;  Sequence 1, Application PC/TUS9604294
;  GENERAL INFORMATION:
;    APPLICANT:  Dixon, Eric P.
;    APPLICANT:  Johnstone, Edward M.
;    APPLICANT:  Little, Sheila P.
```

```
    TITLE OF INVENTION:  AMYLOID PRECURSOR PROTEIN PROTEASE AND
    TITLE OF INVENTION:  RELATED NUCLEIC ACIDS
    NUMBER OF SEQUENCES:  3
    CORRESPONDENCE ADDRESS:
        ADDRESSEE:  Eli Lilly and Company
        STREET:  Lilly Corporate Center
        CITY:  Indianapolis
        STATE:  Indiana
        COUNTRY:  United States of America
        ZIP:  46285
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER:  IBM PC compatible
        OPERATING SYSTEM:  PC-DOS/MS-DOS
        SOFTWARE:  PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER:  PCT/US96/04294
        FILING DATE:
        CLASSIFICATION:
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  US 08/416,257
        FILING DATE:  04-APR-1995
    ATTORNEY/AGENT INFORMATION:
        NAME:  Blalock, Donna K.
        REGISTRATION NUMBER:  38,082
        REFERENCE/DOCKET NUMBER:  X9239
    TELECOMMUNICATION INFORMATION:
        TELEPHONE:  317-277-1090
        TELEFAX:  317-276-3861
    INFORMATION FOR SEQ ID NO:  1:
    SEQUENCE CHARACTERISTICS:
        LENGTH:  1089 base pairs
        TYPE:  nucleic acid
        STRANDEDNESS:  single
        TOPOLOGY:  linear
    MOLECULE TYPE:  cDNA
PCT-US96-04294-1
```

```
Query Match        13.2%;  Score 154.6;  DB 5;  Length 1089;
Best Local Similarity  55.1%;  Pred. No. 9.7e-33;
Matches 351;  Conservative    0;  Mismatches  271;  Indels   15;  Gaps   2;

Qy     2 CGCAGCCCTGCAGGGGCACTGGCATCATGGAAAAACGAATTGTTCTGCTCGGCGTCCTGG 61
           | |  || ||||||| || ||||  ||| || ||||||||||||||||||||||||||||
Db   855 CCCACCCATGGCAGGTGGCCCCTGCTCCAGTGGCAATCAGCTCCACTGCGGGAGGCGTCCTGG 796

Qy    62 TGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTTCCAGAACTCCTACACCATCGGGC 121
           | | |||||||| |||||  ||||| | |||||||||||||||||||||||||||||||
Db   795 TCAATGAGCGGCTGGGGTGCTCACTGCGCGCCCACTGCCAAGATGAATGAGTACACCGTGCACC 736

Qy   122 TGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGAGGCCAGCC 181
           ||||  |||| |||||| |||||||| |||||||||||||       ||||||||||||
Db   735 TGGGCAGTGATACGCTGGGCGACAGGAGAG------CTCAGAGGATCAAGGCCTCGA 685

Qy   182 TCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTCCGTAACGACCTCATGCTCATCA 241
           || |  ||| |||| ||||| || | ||| | ||||||||||||| ||||||||||||||
Db   684 AGTCATTCCGCCACCCGGCTACTCCACACAGACCCATGTTAATGACCTCATGCTCGTGA 625

Qy   242 AGTTGGACGAATCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCCCAGT 301
           |||  || |||   ||| | |||||| || ||   ||||||||| ||| ||||||||||
Db   624 AGCTCAATAGCCAGGCAGCCTGTCTCATCCATGGTGAAGAAAGTCAGGCTGCCCTCCGGCT 565

Qy   302 GCCCTACCGCGGGGAACTCTTGCCTCGTNTCTGGCCGGTGGGGTCTGCTGGCGAAC------G 355
           || ||  |||||  ||||| |||| |||| ||||||||||| ||| | |||||  ||
Db   564 GCGAACCCCCTGGAACCACCTGTACTGTCTCCCGGCACTACCACCAGCCCCAGATG 505

Qy   356 GCAGAATGCCTACCGTGCTCACTGCGTGAACGTGTCGGTGGTGTCTGAGGANGTCTGCA 415
           ||||| ||||  |||| ||||| | || |||||| |  ||||| || |   || |||||
Db   504 TGACCTTCCCTCTGACCTCATGTCGGTGATGTCAAGCTCATCTTCCCCCAGGACTGCA 445

Qy   416 GTAAGCTCATGACCCGCGTGTACCACCCCAGCATGTTCTGCGCGCCGGCGGAGGGCAAGACC 475
           |||  ||  ||  |||   |||| || ||| ||||| ||| |  | |  ||| | |||
```

```
Qy  122 TGGGCCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGGAGCCAGATGTGGAGGCCAGCC 181
        |||||  ||   |||  |  |||  |  |||  |   ||
Db  735 TGGGCAGTGATACGCTGGGCGACCAGGAGAG-------CTCAGAGGATCAAGGCCTCGA 685

Qy  182 TCTCCGTACGGCACCCAGATACAACAGACTCTTGCTGCGCTAACGACCTCATGCTCATCA 241
        || ||||| ||| | |||  || ||| |||| |||||| ||||| ||||||||||| |||
Db  684 AGTCATTCCGCCACCCGGCTACTCCACCAGACCCATGTTAATGACCTCATGCTCGTGA 625

Qy  242 AGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCCGAGCCATCAGCATTGCTTCTTGCAGT 301
        |||||| ||||| | || |   |||| ||||||||||||| |||  |   |||||||| ||
Db  624 AGTCAATAGCCAGGCCAGGCTGTCATCCATGGTGAGAAAGTCAGGCTGCCCTCCCGCT 565

Qy  302 GCCCTACCGCGGGGGAACTCTTGCCTCGTNTCTGGCGTGGGTCTGCTGCTGGCGAAC------G 355
        |||  | |  | ||||||| | || || | | ||| |||| | || | |  |||
Db  564 GCGAACCCCCTGGAACCACCTGTACTGTCTCCCGGCTGGGGCCACTACCACGAGCCCAGATG 505

Qy  356 GCAGAATGCCTACCGTGCTGCGTGGAACGTGTCGGTGGTGTCTGAGGAGNTCTGCA 415
        |||| |  ||||   | ||||| || |||| || ||  |  ||| |  |  ||||
Db  504 TGACCTTCCCTCTGACCTCATGTGCGTGGAGTGCAAGTCAATCTCCCCCGCGGAGGACTGCA 445

Qy  416 GTAAGCTCATGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGCGGAGGGCAAGACC 475
        || |||| ||| |||  |  ||  || | |||   |  | | || |  |
Db  444 CGAAGGTTTACAAGGACTTACTGGAAAATTCCATGTGTGTGCGCTGCTGCATCCCGACTCCA 385

Qy  476 AGAAGGACTCCTGCAACGGTGACTCTGGGGGCCCTGATCTGCAACGGTACTTGCAGG 535
        ||| | ||  ||| |   ||| |  | | ||| ||  |
Db  384 AGAAAAACGCCTGCAATGTGACTGACTCAGGGGGACCGTTGGTGTGCAGAGGTACCCTGCAAG 325

Qy  536 GCCTTGGTGTCTTTCGGAAAAGCCCCGTGTGCCAACTGGCGTGCCAGGTGTCTCACCA 595
        |||  ||| |||  ||| |||| |   || || |  ||| |||| |   ||
Db  324 GTCTGGTGTCCTGGGGAACTTCCCTTGCGCCGTGCCGCCAACCCAATGACCCAGGAGTCTCACACTC 265

Qy  596 ACCCTGCAAATTCACTGAGTGGATAGAGAAAACCGT 632
        |||| | ||  ||| | |  || |||||| || ||
Db  264 AAGTCTGCAAGTTCACCAAGTGGATAAATGACACCAT 228
```

RESULT   8
US-08-557-146-1
; Sequence 1, Application US/08557146
; Patent No. 5684290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 986 base pairs

```
Db  444 CGAAGGTTTACAAGGACTTACTGAAATTCCATGCTGTGCGCTGGCATCCCCGACTCCA 385
Qy  476 AGAAGGACTCCTGCAACGGTGACTCTGGGGGCCCTGATCTGCAACGGTACTTGCAGG 535
Db  384 AGAAAAACGCCTGCAATGTGACTGACTCAGGGGGACCGTTGGTGTGCAGAGGTACCCTGCAAG 325
Qy  536 GCCTTGGTGTCTTTCGGAAAAGCCCCGTGTGCCAACTGGCGTGCCAGGTGTCTCACCA 595
Db  324 GTCTGGTGTCCTGGGGAACTTCCCTTGCGCCGTGCCGCCAACCCAATGACCCAGGAGTCTCACACTC 265
Qy  596 ACCCTGCAAATTCACTGAGTGGATAGAGAAAACCGT 632
Db  264 AAGTCTGCAAGTTCACCAAGTGGATAAATGACACCAT 228
```

RESULT   7
PCT-US96-04294-3/c
; Sequence 3, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1089 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
PCT-US96-04294-3

Query Match   13.2%;   Score 154.6;   DB 5;   Length 1089;
Best Local Similarity   55.1%;   Pred No. 9.7e-33;
Matches   351;   Conservative   0;   Mismatches   271;   Indels   15;   Gaps   2;

```
Qy   2 CGGAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTCGGGGCGTCCTGG 61
       | ||| ||||||||  ||| | ||| ||| ||||  |  ||  | || | ||| ||||||
Db 855 CCCACCCATGGCAGGTGGCCCCTGCCTCAGTGGCAATCAGCTCCACTGGCGGAGGGCTCCTGG 796

Qy  62 TGCATCCGCAGTGGGTGCTGTCAGCCCGCACACTGTTCCAGAACTGTTACACCATCGGGC 121
       |||| | | |||||||||| || |  |   || | |  |||||  |  |   || ||||
Db 795 TCAATGAGCGCTGGGTGCTCACTCGCCGCCCACTGCCAAGAGATGAATGAGTACACCGTGCACC 736
```

```
;       TYPE: nucleic acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;    MOLECULE TYPE: cDNA
;    HYPOTHETICAL: NO
;    ANTI-SENSE: NO
;    ORIGINAL SOURCE:
;       ORGANISM: Homo sapiens
;    FEATURE:
;       NAME/KEY: CDS
;       LOCATION: 25..786
;    FEATURE:
;       NAME/KEY: sig_peptide
;       LOCATION: 25..90
;    FEATURE:
;       NAME/KEY: mat_peptide
;       LOCATION: 112..783
US-08-557-146-1
```

```
;    GENERAL INFORMATION:
;       APPLICANT: Kotin, Robert M.
;       APPLICANT: Berns, Kenneth I.
;       APPLICANT: Linden, Ralph M.
;    TITLE OF INVENTION: Human Adeno-Associated Virus Integration
;    TITLE OF INVENTION: Site DNA and Uses Thereof
;    NUMBER OF SEQUENCES: 23
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;       STREET: Two Militia Drive
;       CITY: Lexington
;       STATE: MA
;       COUNTRY: USA
;       ZIP: 02173
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/308,949A
;       FILING DATE: September 20, 1994
;       CLASSIFICATION: 435
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 07/947,127
;       FILING DATE: September 27, 1992
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Carroll, Alice O.
;       REGISTRATION NUMBER: 33,542
;       REFERENCE/DOCKET NUMBER: ACC92-10F
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 617-861-6240
;       TELEFAX: 617-861-9540
;    INFORMATION FOR SEQ ID NO:  1:
;       SEQUENCE CHARACTERISTICS:
;          LENGTH: 4060 base pairs
;          TYPE: nucleic acid
;          STRANDEDNESS: single
;          TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1
```

```
Query Match            13.1%;    Score 153;  DB 4;   Length 986;
Best Local Similarity  54.9%;    Pred. No. 2.5e-32;
Matches  350;  Conservative   0;  Mismatches  272;  Indels  15;  Gaps   2;

Qy      2 CGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTCGGCGTCCTGG   61
            |   ||| ||||||| |||  ||| |  | |||||  ||| || ||| || ||||||
Db    143 CCCACCCATGGCAGGTGGCCCTGCTCAGTGGCAATCAGCTCCACTGCGGAGGCGTCCTGG  202

Qy     62 TGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTTCCAGAACTCCTACACCATCGGGC  121
           |  || |||||| |||||| || |       |||      ||| |  ||||||
Db    203 TCAATGAGGCGCTGGGTGCTCACTGCCGCCCACTGCAAGATGAATGAGTACACCGTGCACC  262

Qy    122 TGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCC  181
            || |   | |  ||  |  ||||     |||||         |||| |
Db    263 TGGGCAGTGATACGTCTGGGCGACAGGAGAG------CTCAGAGGATCAAGGCCTCGA  313

Qy    182 TCTCCGTACGGCACCCAGAGTACAAACAGACTCTTGCTCGCTAACGACCTCATGCTCATCA  241
            ||  |   ||||| ||||||| |  |      |   |||| ||| |  |||| |||
Db    314 AGTCATTCCGCCACCCCGGCTACTCCACACAGACCCATGTTAATGACCTCATGCTCGTGA  373

Qy    242 AGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCCGAGCATCAGCATTGCTTCGCAGT  301
            ||        | |   || |    ||    |      ||||| |   |  || ||||
Db    374 AGCTCAATAGCCAGGCCAGGCGTGTCATCCATGGTGAAGAAAGTCAGGCTGCCTGCCCCCGCT  433

Qy    302 GCCCTACCGCGGGGAACTCTTGCCTCGTGTNTCTGGCTGGGGTCGTCGCTGGCA------ACG  355
            || |   |||| |||||| ||||| || |    | |||||| |  ||| |   |      |||
Db    434 GCGAACCCCCTGGAACCACCTGTACTGTCTCCGGCTGGGGCACTACCACGAGCCCAGATG  493

Qy    356 GCAGAATGCCTACCGTGCTGCACTGCCTGACGTGTCGGTGGTGTCTGAGGANGTCTGCA  415
            ||    |  | |         || ||  | |||| ||   |  |||  |
Db    494 TGACCTTTCCCTCTGACCTCATGTGCCTGACGTTCAAGCTCATCTCCCCCAGACTGCA  553

Qy    416 GTAAGCTCTATGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGCGAGGGCAAGACC  475
            ||  | || | |  |   |       ||       ||| ||||
Db    554 CGAAGGTTTACAAGGACTTACTGGAAAATTCCATGCTGTCGCCGTGCTGGCATCCCGGACTCCA  613

Qy    476 AGAAGGAGACTCCTGCACACGGTGACCGTGACTCTGGGGGCCCCTGATCTGCAACGGGTACTTGCAGG  535
            ||    |  | |  | ||   |||    ||    |||||||||   |  | |||| ||| || ||||
Db    614 AGAAAAACGCCTGCAATGGTGACTCAGGGGGACCGTTGGTGTGCAGAGTACCGGCAAG  673

Qy    536 GCCTTGTGTCTTTTCGGAAAAGCCCCGTGTGCCCAACTTGGCGTGCCAGGTGTCTACACCA  595
            || ||||||  |      ||||| ||   |  |    ||| | |||||| |
Db    674 GTCTGGTGTCCTGGGGAACTTTCCCTTGCGCGCCAACCAATGACCCCAGGAGTCTCACACTC  733

Qy    596 ACCTCTGCAAATTCACTGAGTGGATTAGAGAAAAACCGT  632
            ||| | |  ||   | |  |  |   |   ||| |
Db    734 AAGTGTGCAAGTTCACCAAGTGGATTAAATGACACCAT  770
```

```
RESULT  9
US-08-308-949A-1/c
; Sequence 1, Application US/08308949A
; Patent No. 5580703
```

```
Query Match             12.2%;    Score 142.2;  DB 1;   Length 4060;
Best Local Similarity   71.6%;    Pred. No. 3.4e-29;
Matches  240;  Conservative    0;  Mismatches  81;  Indels  14;  Gaps   4;

Qy    718 CCAGCCCCTCCTCCCCTCCCCAGGGCCAGGAGTCCAGGCCCCACCCTCCCTCCTCCTCAAACCA  777
            || |     |  || |||||||||| ||||||| |||||||||   | |
Db   4000 CCCAACCCCTCCCCCATTCAACCCAGGAGGCCAGGCCCCTTCCGCCCCTTCAGATGAA  3941

Qy    778 AGGGTACAGATCCCCAGCCCCTCTCCCTCCCCATTCCCCATTCAGACCCAGACCCCCAGCCCCT  837
            ||||| |   | |  || |||||| ||||| |||||||| ||||  ||||| || ||||||||
Db   3940 GGAGTCCAGGCCCCAGCCCTCCCAGTCCCCATTCAGACCCAGCAGGGGTCCAG---GCCCAGCCCCG  3884

Qy    838 CNTCCNTCAGACCCAGGAGTCCAG-------CCCTCCTCCNTCAGACAGGCAGGAGTCCA  889
            |   |  | || |  |||| ||||       ||||||| | |||||| || || || ||
Db   3883 CCTCCCTAAGACCCAGGAAGTCCAGGCCCCCGACCCCTCCCTCAGACCCACGAGTCCA  3824

Qy    890 GACCCCCAGCCCNCNTCNTCCCTCCTCCTCGGACGTGCAGGCCCCAGGGGTGCAGGCCCCCCAACCCTCNTCCNTC  949
            | |||||||  |  |  | ||||||| ||||||||||| |||| || |||  | ||||||| ||||||||| ||||
Db   3823 G--GCCCCAGCCCCTCCTCCTCGGACCCCAGGAGTCCAGGCCCCAGCTCCTCCAGTCCCTCCCACCCTC  3766

Qy    950 AGAGTCAGGAGTCCAAGCCCCCAACCCCTCGTTCCCAGACCCCAGGAGGTNCAGGTCCCAG  1009
            ||| ||| || |  |||||||  || || |  ||||||| ||||||||| |||| ||||||
Db   3765 AGACCCAGGAGTCC-AGGCCCCAGCCCCTCCTCCCCTCGGAGACCCAGGAGTCCAGGCCCCCAG  3707

Qy   1010 CCCCTCCTCCCTCAGACCCCAGCCGGTCCAATGCCAC  1044
            |||||||||  |||||||  |  ||      |   ||
Db   3706 CCCCTCCTCTCCTCAAACCCAGGAGGACCCCAGGCCCCCCCC  3672
```

RESULT 10

US-08-361-395-2
; Sequence 2, Application US/08361395
; Patent No. 5733768
; GENERAL INFORMATION:
;   APPLICANT: Dixon, Eric P.
;   APPLICANT: Johnstone, Edward M.
;   APPLICANT: Little, Sheila P.
;   APPLICANT: No. 5733768ris, Franklin H.
;   TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Eli Lilly and Company
;   STREET: Lilly Corporate Center/Patent Division
;   CITY: Indianapolis
;   STATE: Indiana
;   COUNTRY: United States
;   ZIP: 46285
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/361,395
;   FILING DATE: 22-DEC-1994
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Blalock, Donna K.
;   REGISTRATION NUMBER: 38,082
;   REFERENCE/DOCKET NUMBER: X8350A
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 317-277-1090
;   TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 732 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-361-395-2

Query Match   12.0%; Score 140; DB 3; Length 732;
Best Local Similarity 52.1%; Pred. No. 6;9e-29;
Matches 333; Conservative 0; Mismatches 303; Indels 3; Gaps 1;

```
Qy      2 CGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTCGGGCGTCCTGG 61
Db     95 CTCACCCCTACCAAGCTGCCCTCTACACCTCGGGCCACTTGCTCTGTGGTGGGGTCCTTA 154

Qy     62 TGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTTCCAGAACTCCTACCACCATCGGGC 121
Db    155 TCCATCCACTGTGGGTCCTCACAGCTGCCCACTGCCAAAAAACGAATCTTCAGGTCTTCC 214

Qy    122 TGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGGATGGTGAGGCCAGCC 181
Db    215 TGGGGAAGCATAACCTTCGGCAAAGGGAGAGTTCCCAGGCAGAGTTCGTTG---TCC 271

Qy    182 TCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTCCTCGCTAACGACCTCATCGTCATCA 241
Db    272 GGGCTGTGATCCACCCTGACTATGATGCCGCGAGCCATGATGCCTGTTGC 331

Qy    242 AGTTGGACGAATCCGTCCGAGTCTGACACCATCCGAGCAGCATCAGCCATTGCTTCGCAGT 301
Db    332 GCCTGGCACGCCCAGCCAAACTCTCTGAACTCATCCAGCCCTTCCCCTGGAGAGGGACT 391

Qy    302 GCCCTACCCGCGGGAACTCTTGCCCTCGTNTCTGGCTGGGCTCTGGCGAACGGCAGAA 361
Db    392 GCTCAGCCAACACCACCAGCTGCCACATCCTGGGCTGGGCAAGAACAGCAGATGGTGATT 451

Qy    362 TGCCTACCGTGCTGCTGCACTGCGTGAACGTGCTGGTTGTCTGCGAGANGTCTCAGTAAGC 421
```

Db   452 TCCCTGACACCATCCAGTGTGCATACATCCACCTGTGTCCCGTGTCCCGTGAGGAGTGTGAGCATG 511
Qy   422 TCTATGACCCGCTGACCACCCCAGCATGTTCTCGCGCGGAGGGCAAGACCAGAAGG 481
Db   512 CCTACCCTGGCCAGATCACCCCAGAACATGTTGTGTCTGTGGGATGAGAAGTACGGGAGG 571
Qy   482 ACTCCTGCAACGGTGACTCTGACTGTGGGGGCGCCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 541
Db   572 ATTCCTGCCAGGTGATTCTGGGGGTCTGGGGTGTATGTGGAGACCACCTCCGAGGCCTTG 631
Qy   542 TGTCTTTCGGAAAAGCCCGTGTGGCCAACTTGGCGTGCCAGTGTCTACACCAACCTCT 601
Db   632 TGTCATGGGGTAACATCCCCTGGATCAAAGGAGGAGCCAGGAGTCTACACCAACGTCT 691
Qy   602 GCAAATTCACTGAGTGGATAGAGAAAAACGTCCAGNCCA 640
Db   692 GCAGATACACGAACTGGATCCAAAAAACCATTCAGGCCA 730

RESULT 11
US-08-790-137-2
; Sequence 2, Application US/08790137
; Patent No. 5840871
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Goll, Surya K.
;   TITLE OF INVENTION: A NOVEL PROSTATE-ASSOCIATED
;   TITLE OF INVENTION: KALLIKREIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/790,137
;   FILING DATE: Filed Herewith
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0195 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 833 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-790-137-2

Query Match   10.7%; Score 125.4; DB 4; Length 833;
Best Local Similarity 52.6%; Pred. No. 5.9e-25;
Matches 348; Conservative 0; Mismatches 292; Indels 21; Gaps 3;

```
Qy      2 CGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTCGGGCGTCCTGG 61
Db    130 CCCAGCCCTGGCAGGCGGCTCTGTACCAGAAGACGCGGCTACTCTGTGGGGCGACGNTCA 189

Qy     62 TGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTTCCAGAACTCCTACCACCATCGGGC 121
```

```
Db  190 TNGCCCCAGATGGTTCCTGACAGCAGCCCACTGCTGCCTNAAGCCCGCTACATAGTTCACC 249
          ||    |||||  ||| ||||| |||||  |  ||||     |  ||  |||||
Qy  122 TGGGCCCTGCACAGTCTGAGGCCAGAGAGCCAGGAGCCAGATGGTGAGGGCCAGCC 181
          |||| |||| ||||   |  |  |||||||||| |||  |            ||||
Db  250 TGGGCAGCACAACCTCCAGAAGGAGGAGGGCTGTGAG---CAGACCCGGACAGCCACTG 306
```

```
Qy  182 TCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTTCGCTAACGACC-------- 230
          ||| | ||| |    ||||||    ||| ||   ||||
Db  307 AGTCCTTCCCCACCCCGGCCTCAACAACAGCCTCCCAACAAAGACCACCGCAATGACA 366
```

```
Qy  230 TCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCA 289
          |||  |   |||  | |  ||     ||| |  ||    ||  ||    ||||||
Db  367 TCATGCTGGTGAAGATGCATCGCCAGATCTCCATCACCTGGCGTGTGCGACCCCTCACCC 426
```

```
Qy  290 TTGCTTCGCAGTGCCCTACCACCGCGGGGAACTTGCCTCGTNTCTGCCTGGGTCTTGCTGG 349
          | |   ||    | ||||  |  ||||||   || ||| |  |||||  |||||||||||
Db  427 TCTCCTCACGCTGTGTCACTGCTGCCTCATTTCCGGCTGGGCAGCACGT 486
```

```
Qy  350 CGAACGGC------AGAATGCCTACGGTGCTGCACTGCGTGAACGTGTCGGTGGTGTCTG 403
          | |            |     | |  |||| | | ||| |  |||  || | |   |
Db  487 CCAGCCCCCAGTACGCCTGCCTCCACACCTTGCGATGCGCCAACATCACCATCATTGAGC 546
```

```
Qy  404 AGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATGTTCTGCCCGCGCG 463
          |||| |   |  |  |     ||     ||  |      ||||    ||   | |||||
Db  547 ACCAGAAGTGTGAGAACGTGTACCCCGCAACATCACAGACACCACCATGGTGTCGCCAGCG 606
```

```
Qy  464 GAGGGCAAGACCAGAGGACTCCTGCAACAGGTGACCTGGGGGGCCCCCTGATCTGCAACG 523
          |||||||||||||||||||||||     |||  | | |||||  |||||||   |   |
Db  607 TGCAGGAAGGGGGCAAGGACTCCTGCCAGGTGACTCCCGGGGCCCCTGGTCTGTAACC 666
```

```
Qy  524 GGTACTTGCAGGGCCTTGTGTCTTCGGAAAAGCCCGTGTGGCCAACTTGGCGTGCCAG 583
          ||  |   || | |  ||| | ||    |||||  |  ||||||   |||||||||
Db  667 AGTCTCTTCAAGGCATTATCTCCTGGGGCCAGGATCCGTCGTGCGATCCACCCGAAAGCCTG 726
```

```
Qy  584 GTGTCTACACCACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCGTCCAGNCCAGTT 643
          |||||||    ||| |||    |  | |  || || |   ||| |||||  ||| |  ||
Db  727 GTGTCTACACGAAAGTCTGCAAATATGTGGACTGGACTCGAGGACGATGAAGAACAATT 786
```

```
Qy  644 A 644
          |
Db  787 A 787
```

RESULT 12
US-08-684-862-11
; Sequence 11, Application US/08684862
; Patent No. 5759541
GENERAL INFORMATION:
  APPLICANT: Bach, Alfred
  APPLICANT: Hillen, Heinz
  APPLICANT: Bialojan, Siegfried
  TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
  TITLE OF INVENTION: Thereof
  NUMBER OF SEQUENCES: 14
  CORRESPONDENCE ADDRESS:
     ADDRESSEE: Keil & Weinkauf
     STREET: 1101 Connecticut Avenue
     CITY: Washington
     STATE: D.C.
     COUNTRY: USA
     ZIP: 20036
  COMPUTER READABLE FORM:
     MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
     COMPUTER: IBM AT-compatible, 80286 processor
     OPERATING SYSTEM: MS-DOS version 5.0
     SOFTWARE: WordPerfect version 5.1
  CURRENT APPLICATION DATA:
     APPLICATION NUMBER: US/08/684,862
     FILING DATE:
     CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
     APPLICATION NUMBER: US/08/361,705

  FILING DATE:
  PRIOR APPLICATION DATA:
     APPLICATION NUMBER: 07/966,040
     FILING DATE: 30-DEC-1992
     APPLICATION NUMBER: PCT/EP91/01361
     FILING DATE: 19-JUL-1991
INFORMATION FOR SEQ ID NO: 11:
  SEQUENCE CHARACTERISTICS:
     LENGTH: 957 base pairs
     TYPE: nucleic acid
     STRANDEDNESS: single
     TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  ORIGINAL SOURCE:
     ORGANISM: Agkistrodon rhodostoma
  FEATURE:
     LOCATION: 210 to 911
     OTHER INFORMATION: the coding region shown in (2)(ix)(B)
     OTHER INFORMATION: codes for the protein of SEQ ID NO: 6
US-08-684-862-11

Query Match   8.2%; Score 95.8; DB 3; Length 957;
Best Local Similarity 49.9%; Pred. No. 5.3e-17;
Matches 297; Conservative   0; Mismatches 289; Indels   9; Gaps   2;

```
Qy   40 TTGTTCTGTCTCGGCGTCTCGTGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTTTC 99
          || ||||| |    ||||  || | |   || |  | |||| |||    ||| |||||
Db  279 TTTGTCTGCGCTGGGACTTTGATCCACCCAGAATGCGGTGGTCACTGCTGCACACTGCGAG 338
```

```
Qy  100 CAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGG 159
          |     | || |  |  ||    | ||    ||  |  |||||||  || || |||||
Db  339 AGTACGGATCTCAAGATGAAGTTTGGTATGCATAGCAAAAGGTACAAAAATGAGGATGAG 398
```

```
Qy  160 AGCCAGATGGTGGAGGCCAGCCTCTCGTACGGCACCCAGTACTACAACAGACTCTTGCTC 219
          |||   ||   |  |    |  ||    |||   ||||   |   |||  || ||| ||
Db  399 CAGACAAGAAACGCAAAGGAAAGTTCATTTGTCCCAATAAGATAGAAGTAACGATGACTG 458
```

```
Qy  220 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCGTGACCACCATCCGG 279
          |   ||  ||   | || ||| ||     ||| |  |||| |  || ||   |||  ||
Db  459 GACAAGGACATTATGTTGATCAAGCTGAACCATCCTGTTAGCAATAGTGAACAACATCGCG 518
```

```
Qy  280 AGCATCAGCATTGCTTCGCAGTGCCCTACCACCGCGGGGAACTCTTGCCTCGTNTCTGCGTGG 339
          || |   |||  |  ||  | |||  ||   |||  |  ||   ||||  ||  || |
Db  519 CCTCTCCAGCTTGCCTTCCAGCCCTCCAGTGTGGGCTCATTTGCCATATTATGGGATGG 578
```

```
Qy  340 GGTCTGTCGGCGAACGGCAGAATGCCTACC------GTGCTGCACTGCGTGAACGTGTCG 393
          || |          | |||  | || ||      ||||||| | | | || |  |||
Db  579 GGCTCAATCACACCTGTTAAAGTGACTTTCCCCGATGTCCCTGCTAACATTAAC 638
```

```
Qy  394 GTGGTCTGTGAGGANGTCTGCAGTAAGCTCTA---TGACCCGCTGTACCACCCCAGCATG 450
          ||  | || ||   | ||| || || || ||   |||  |  ||  |  || ||||||
Db  639 CTACTCGATGATGCAGAGTGTCATGCAGTTGGCAGGTCTTGGCAGAATACAGAACA 698
```

```
Qy  451 TTCTGCCCGCGCGGAGGGCAAGACCAGAGGACTCCTGCAACAGGTGACCTGACTCGGGGGGCCC 510
          |||||||  ||||  |  |||  |||  |  |  |   |||| |||  || | ||||  ||
Db  699 TTGTGTGCAGTATCGTGCAAGGAGGCAAAGATACATGTATGACTCTGGAGGACCT 758
```

```
Qy  511 CTGATCTGCAACGGGTACTTGCGAGGGCCTTGCTGGCCGCCAACCTCTGCAAATTCACTGAGTGGATAGAGA 570
          || |  | || ||  || ||   |||  |   |||||   ||    |  |||| |||  ||||||||
Db  759 CTCATCTGTAATGAACAAGTCCAGGGCATTGTATCTTATGGGGCGCCATCCTTGTGCCAA 818
```

```
Qy  571 CTTGGCGTGCCAGTGTCTCACACCACCAACCTCTGCAAATTCACTGAGTGGATAGAGA 625
          || |    | |  ||| ||||  |  ||||     | || |  |||||||   | |
Db  819 CCTCTTAAGCCTGGTATCTCACACCAGGCTCCAGATTAATAATGACTGGATCAACA 873
```

RESULT 13
US-08-358-782D-14
; Sequence 14, Application US/08358782D
; Patent No. 5674682
; GENERAL INFORMATION:
     APPLICANT: Croce, Carlo

2;

```
Db  545 TGGACCTCCATGTTATTCCAATGACGTGTGCGCAAGTTCACCCTCAGGAGGTGACCA 604
            ||||     || |||     ||| ||  ||       || |    || |||
Qy  443 CCAGCATGTTCGCCGCGGCGGGGGAGGCAAGACCACCAGAAGGACTCCTGCAACGGTGACTCTG 502
            |||||| || |||    || | ||  |||
Db  605 AGTTCATGCTGTGTCGCTGGACGCTGGACAGGGGGCAAAAGCACCTGCTCGGGTGATTCTG 664
            |||| |||||||||
Qy  503 GGGGGCCCTGATCTGCAACGGTACTTGCAGGGCCTTGTGTCTTTCGGAAAAGCCCGT 562
            |||| ||||| |||| || |||           ||| |    || |||
Db  665 GGGGCCCACTTGTCTGTAATGGTGTGCTTCAAGGTATCACGTCATGGGGCCAGTGAACCAT 724
            |||| |  |||||
Qy  563 GTGGCCAACTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAATTCACTGAGTGGATAG 622
            || ||||     |||   ||             ||||||
Db  725 GTGCCCTGCCCGAAAGCCTTCCCTGTACACCAAGGTGGTGCATTACCGGAAGTGGATCA 784
            || |||||  |||
Qy  623 AGAAAACCGTCCAGNCCA 640
            || |||||| | |||
Db  785 AGGACACCATCGTGGCCA 802
```

RESULT 14
US-08-358-782D-13
; Sequence 13, Application US/08358782D
; Patent No. 5674682
; GENERAL INFORMATION:
;   APPLICANT: Croce, Carlo
;   APPLICANT: Gomella, Leonard
;   APPLICANT: Mulholland, S. Grant
;   APPLICANT: Moreno, Jose
;   APPLICANT: Fischer, Rainer
;   TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
;   STREET: One Liberty Place 46th. Floor
;   CITY: Philadelphia
;   STATE: PA
;   ZIP: 19103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/358,782D
;   FILING DATE: 15-DEC-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Beardell, Lori Y.
;   REGISTRATION NUMBER: 34,293
;   REFERENCE/DOCKET NUMBER: TJU-1327
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-568-3100
;   TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 992 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
US-08-358-782D-13

Query Match   7.2%;   Score 84.2; DB 2; Length 992;
Best Local Similarity 49.4%; Pred. No. 6.9e-14;
Matches 335; Conservative 0; Mismatches 301; Indels 42; Gaps 3;

```
Qy    2 CGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAAACGAATTGTTCTGCTCGGGCGTCCTGG 61
            | ||||||||||||                           ||| ||||| || ||||
Db  147 CCCAACCCCTGGCAGGTGCTTGTAGCCTCTCTGGGCGCATGAGAAAAACGAATTGTTCTGCTCGGGTCGGGGTCTCGG 206
```

APPLICANT: Gomella, Leonard
APPLICANT: Mulholland, S. Grant
APPLICANT: Moreno, Jose
APPLICANT: Fischer, Rainer
TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
  STREET: One Liberty Place 46th. Floor
  CITY: Philadelphia
  STATE: PA
  ZIP: 19103
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/358,782D
  FILING DATE: 15-DEC-1994
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: Beardell, Lori Y.
  REGISTRATION NUMBER: 34,293
  REFERENCE/DOCKET NUMBER: TJU-1327
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 215-568-3100
  TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
  LENGTH: 1462 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
US-08-358-782D-14

Query Match   7.4%;   Score 85.8; DB 2; Length 1462;
Best Local Similarity 49.6%; Pred. No. 3e-14;
Matches 336; Conservative 0; Mismatches 300; Indels 42; Gaps 3;

```
Qy    2 CGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTCGGGCGTCCTGG 61
            | ||||||||||||                      ||| ||||| ||  ||||
Db  128 CCCAACCCCTGGCAGTGCTTGTGGGCCTCTCGGCGAGGGCAGTCTCGGCGGGTGTTCTGG 187

Qy   62 TGCATCCGCACTGGGTGCTGTCAGCCGCACACTGTTCCAGAACTCCTACACCATCGGGC 121
            ||||  || |||||||| || ||||||||||| || ||||
Db  188 TGCACCCCCAGTGGGTGCTCCCACAGCTGCCCACTCGCATCAGGAACAAAAGCGTGATCTTGC 247

Qy  122 TGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCC 181
            |||| |||||||| ||          |||||| || ||| |||||
Db  248 TGGGTCGGCACAGCCGTGTTCATCCTGAGACACAG--GCCAGGTATTTCAGGTCAGCC 304

Qy  182 TCTCCGTACGGCACCCAGAGTACAACAGACTCTTGCTCCGCTA----------- 224
            |||   ||| ||||   |||||   || ||
Db  305 ACAGCTTCCCACACCCGCTTACGATATGAGCCTCCTGAAGAATCGATTCCTCAGGCCAG 364

Qy  224 ----------ACGACCTCATGTCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTG 268
                       |||||||||  || |||| ||||
Db  365 GTGATGACTCCAGCCACGACCTCATGTCGTCTCCGCCTGTGCAGAGCCTGCCGAGCTCACCG 424

Qy  269 ACACCATCCGGAGCATCAGCATTGCTTCGGCAGTGCCCTCGCCGGGGAACTCTTGCCTCG 328
            || |||| ||  || | |||||||  |||    ||||| | ||
Db  425 ATGCTGTGAAGGTCATGGACCTGCCCACCCAGGAGCCAGCACCTGGGAGCACCTGCTACG 484

Qy  329 TNTCTGGCTGGGG-----TCTGCTGGCGAACGGCAGAATGCCTACCGTCGTCACTGCG 382
            || ||||||||||      |||||||||| ||| |||
Db  485 CCTCAGGCTGGGCCGCAGCCATTGAACCAGAGAGTTCTTGACCCCAAAGAAACTTCAGTGTG 544

Qy  383 TGAACGTGTCGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGGCTGTACCACC 442
```

```
                       FILING DATE:  15-APR-1994
                   ATTORNEY/AGENT INFORMATION:
                       NAME:  White, John P.
                       REGISTRATION NUMBER:  28,678
                       REFERENCE/DOCKET NUMBER:  43677/JPW/TEP
                   TELECOMMUNICATION INFORMATION:
                       TELEPHONE:  (212) 977-9550
                       TELEFAX:    (212) 664-0525
                       TELEX:  422523 COOP UI
                 INFORMATION FOR SEQ ID NO:  1:
                       SEQUENCE CHARACTERISTICS:
                       LENGTH:  1729 base pairs
                       TYPE:  nucleic acid
                       STRANDEDNESS:  single
                       TOPOLOGY:  linear
                       MOLECULE TYPE:  cDNA
                   ORIGINAL SOURCE:
                       ORGANISM:  Homo sapiens
                   FEATURE:
                       NAME/KEY:  CDS
                       LOCATION:  378..1088
                 US-08-844-024-1
```

```
Query Match           7.2%;  Score 84.2;  DB 4;  Length 1729;
Best Local Similarity 49.4%;  Pred. No. 8.5e-14;
Matches  335;  Conservative    0;  Mismatches 301;  Indels  42;  Gaps  3;

Qy    2  CGCAGGCCTGGCAGGCGGCACTGGTCATGGAAAACGAAATGTTCTGCTGGCGGCGTCCTGG  61
         | ||| ||||||| || | | || |  | || |||| |  | ||||||||| | | |||||
Db  409  CCCAACCCTGGCAGGTGCTTGTAGCCTCTCGTGGCAGGGCAGTCTGCGCGGGTGTTCTGG  468

Qy   62  TGCATCGCCAGTGGGTGCTGTCAGCCGCCACACTGTTCCAGAACTCCTACACCATCGGGC  121
         |||| | ||||||||| ||  ||| |  | |||||  | || || | |  | ||  |||
Db  469  TGCACCCCCAGTGGGTCCTCACAGCTACCCACTGCATCAGGAACAAAAGCGTGAATCTTGC  528

Qy  122  TGGGGCTGCACAGTCTTGAGGCGACCAGGAGCCAGGATGGTGGAGGCCAGCC          181
         |||| | |||||| | | | ||  |||| | | | || ||   ||  ||||||
Db  529  TGGGTCGGCACAGCCTGTTTCATCCTGAAGACACAG--GCCAGGTATTTCAGGTCAGCC    585

Qy  182  TCTCCGTACGGCACCCAGTACAACAGACTCTTGCTCGCTA----------------      224
         | | | |  |||| |  |||| |||| |   | |  | |||
Db  586  ACAGTTCCCACACCCGCTCTACGATATGAGCCTCCTGAAGAATCGATTCCTCAGGCCAG   645

Qy  224  ----------------ACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCGAGTCTG   268
                         ||||||||||||||| | |  |  |  |  | | ||| | |||
Db  646  GTGATGACTCCAGCCACGACCTCATGCTCGCCCGCGCTGTCCGCCGCTGCAGCCTCACGG   705

Qy  269  ACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACTCTTGCCTCG   328
         | | |  | | |  || || |||  | |  || |||   | | ||||| | | |  | |
Db  706  ATGCTATGAAGGTCATGGACCTGACCGCCACCGAGGACCAGCCAGCCACCTGCTACG     765

Qy  329  TNTCTGGCTGGG--------TCTGCTGGGCAACGGCAACGGCAGCAGAATGCTTACCGTGCTCACTGCG   382
         | | ||||||||         | || ||| |  | | |  || | | || || || | |  || | |
Db  766  CCTCAGGCTGGGCAGCATTGAACCAGAGGAGTTCTTGACCCCAAAGAAACTTCAGTGTG   825

Qy  383  TGAACGTGTCGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACC   442
         ||| | || | || | | ||  |  | | | || | |  | | | |||  | |  |||
Db  826  TGGACCTCCATGTTATTTCCAATGACGTGTGTGCGGCCAAGTTCACCCTCAGAAGGTGACCA   885

Qy  443  CCAGCATGTTCTGCGCCGCGGAGGCAAGACCAGAGGACTCCTGCAACGGTGACTCTG     502
         || |  | ||| |  | |  |||| | || || || | || || ||  | | | |||
Db  886  AGTTCATGCTGTGTGCTGACGCTGGAGCAGGGGCAAAAGCACCTGCTCGGGTGATTCTG   945

Qy  503  GGGGGCCCCTGATCTGCAACGGACCTTTGCAGGGCCTTGCAGGGGTACTTGCAGGGTCTTTCGGAAAAGCCCCGT   562
         ||||||||  | | || |  || || |  | | | | | | ||| |   | | ||   | | || | |  | ||
Db  946  GGGGGCCCACTTGCTCGTAATGGTGTGCTTCAAGGTATCACGTCATGGGCAGTGAACCAT   1005

Qy  563  GTGGCCAACTTGGCGTGCCAGTGTCTACACCAACCTCGCAAATTCACTGAGTGGATAG    622
         || |||    || |  | ||  | | ||| || || |  |||| || | || |||| |
Db 1006  GTGCCCTGCCCGAAAGGCCTTCCCTGTACACCAAGGTGGTGCATTACCGGAAGTGGATCA   1065

Qy  623  AGAAAACCGTCCAGNCCA   640
```

```
RESULT 15
US-08-844-024-1
; Sequence 1, Application US/08844024
; Patent No. 5840494
GENERAL INFORMATION:
    APPLICANT:  Katz, Aaron E., et al.
    TITLE OF INVENTION:  A Method For Molecular Staging of
    TITLE OF INVENTION:  Prostate Cancer
    NUMBER OF SEQUENCES:  4
    CORRESPONDENCE ADDRESS:
        ADDRESSEE:  Cooper & Dunham
        STREET:  30 Rockefeller Plaza
        CITY:  New York
        STATE:  New York
        COUNTRY:  U.S.A.
        ZIP:  10112
    COMPUTER READABLE FORM:
        MEDIUM TYPE:  Floppy disk
        COMPUTER:  IBM PC compatible
        OPERATING SYSTEM:  PC-DOS/MS-DOS
        SOFTWARE:  PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER:  US/08/844,024
        FILING DATE:  18-APR-1997
        CLASSIFICATION:  435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  08/229,391
```

Wed Sep 29 14:27:32 1999

us-09-030-606-175_1.rni

```
       || | ||| || | |||
Db   1066 AGGACACCATCGTGGCCA 1083
```

Search completed: September 25, 1999, 07:30:09
Job time: 3446 sec

This Page Blank (uspto)

OM of: US-09-030-606-175 to: Issued_Patents_AA:*   out_format : pfs

Date: Sep 25, 1999 4:01 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171616_29813/app_query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-175
Query length: 1167
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

score_list:

sequence	Strd	Orig.	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1 +	1091.00	1595.26	3.6e-82	248		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2 +	530.50	771.08	2.9e-36	253		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2 +	530.50	771.08	2.9e-36	253		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-12 +	516.00	750.92	4.3e-35	225		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-361-395-1 +	470.50	683.23	2.3e-31	244		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-9 +	456.50	663.53	3.2e-30	223		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-9 +	456.50	663.53	3.2e-30	223		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9 +	456.50	663.53	3.2e-30	223		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-9 +	456.50	663.53	3.2e-30	223		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-3 +	456.00	661.37	3.7e-30	258		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-11 +	450.00	653.38	1.1e-29	237		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11 +	450.00	653.38	1.1e-29	237		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-1 +	450.00	653.38	1.1e-29	237		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-5 +	449.00	650.97	1.4e-29	261		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8 +	448.00	650.40	1.6e-29	238		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-06157-10 +	448.00	650.16	1.6e-29	244		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-6 +	448.00	649.50	1.7e-29	261		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-7 +	447.50	650.09	1.8e-29	228		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-7 +	447.50	650.09	1.8e-29	228		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-7 +	447.50	650.09	1.8e-29	228		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-7 +	447.50	650.09	1.8e-29	228		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-10 +	447.00	648.97	2.0e-29	237		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-10 +	447.00	648.97	2.0e-29	237		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-16 +	447.00	648.97	2.0e-29	237		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-844-024-2 +	443.00	643.09	4.2e-29	286		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-628-198-9 +	429.00	620.67	6.1e-28	286		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-9 +	429.00	620.67	6.1e-28	286		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-4 +	436.50	632.55	1.5e-28	262		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-1 +	436.50	632.55	1.5e-28	262		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-3 +	436.50	632.55	1.5e-28	262		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-681-151-4 +	434.50	630.47	2.1e-28	240		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-472-228A-1 +	434.50	630.47	2.1e-28	240		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US98-09303-1 +	434.50	630.47	2.1e-28	240		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-9 +	429.00	620.67	6.1e-28	286		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-4 +	425.00	615.61	1.3e-27	263		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-7 +	418.50	605.40	4.4e-27	281		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-7 +	418.50	605.40	4.4e-27	281		
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-7 +	418.50	605.40	4.4e-27	281		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-8 +	415.50	602.87	7.4e-27	232		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-8 +	415.50	602.87	7.4e-27	232		
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-8 +	415.50	602.87	7.4e-27	232		
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-8 +	415.50	602.87	7.4e-27	232		

/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-10 + 404.00 584.44 6.7e-26 271
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-10 + 404.00 584.44 6.7e-26 271
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-10 + 404.00 584.44 6.7e-26 271

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1

alignment_scores:
     Quality: 1091.00   Length: 214
     Ratio: 5.195   Gaps: 0
Percent Similarity: 98.131   Percent Identity: 94.860

alignment_block:
US-09-030-606-175 x US-08-744-026-1 ..

Align seg 1/1 to: US-08-744-026-1 from: 1 to: 248

        1 GGGCAGCCCTGCCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTC 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       35 SerGlnProTrpGlnAlaLeuValMetGluAsnGluLeuPheCysSe 51

       51 GGGCGGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACTGTTTCC 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 rGlyValLeuValHisProGlnTrpValLeuSerAlaAlaHisCysPheG 68

      101 AGAACTCCTACACCATCGGCTGGGCCTGCAGACTCTTGAGGCCGACCAA 150
          ||||||||||||||||||||||||||||||||||||||||||||||||

```
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE:   (212) 819-8783
;     TELEFAX:     (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 253 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-557-146-2

alignment_scores:
        Quality:     530.50                    Length:      214
        Ratio:         3.255                    Gaps:          3
  Percent Similarity:  76.168      Percent Identity: 47.196

alignment_block:
US-09-030-606-175 x US-08-557-146-2

Align seg 1/1   to:  US-08-557-146-2   from:  1   to: 253

   1 GCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTC  50
     :::::|||||||||||||||:||:::::  ||:::||  |||::|||
  40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysGl  56

  51 GGGCGTCCTGGTCGCATCCCAGTGGTGCGTGCAGCCCGCACACTGTTTCC  100
     :|||||||||||:::   ::||||||||||:::||||||||:||||||  :
  56 yGlyValLeuValAsnGluArgTrpAlaAlaHisCysLysM  73

 101 AGAACTCCTACACCATGGGCGTGGGCCTGCACAGTCTTGAGGCCGACCAA  150
     ||||||||||||||||||::: ||||||||||||||||  ::::|||:::
  73 etAsnGlyTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg  88

 151 GAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA  200
     ::: :  :::|||  ||||||||:||| |||  |||||||
  89 Arg......AlaGlnArgIleLysAlaSerLysSerPheArgHisProGl  103

 201 GTACAACAGACTCTGCTCGCTAACGACCTCATCGCTCATCAAGTTGGACG  250
     :||:::   ::||||||||||||||:::|||||||||||||:::|||
 103 yTyrSerThrGlnThrHisValAsnAspLeuMetLeuValIleLysLeuAsns  120

 251 AATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCCGCAG  300
     :::::||  |||:   ||||:  |||||||||||:: |||  ||:::
 120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg  136

 301 TGCCCTACCGCGGGGAACTCTTGCCTCGTCNTCCGCTGGGGTCTCGCTGGC  350
     ||   ||:::||||||  ||||||||::|||  ||||||||||||||||
 137 CysGluProGlyThrThrCysThrValSerGlyThrTrpGlyThrThrTrTh  153

 351 GAACGGC.....AGAATGCCTACCGTCGCTGCCACTGGCCGTGAACGTGTCGG  394
     :::    :  :::||||:::  |||   |||||:::||:||::||:
 153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL  170

 395 TGGTGTGTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCC  444
     |::::|||  ||||||||||||||::|||||  ||||||::   ::::
 170 euIleSerProGlnAspCysThrLysValTyrLysAspLeuLeuGluAsn  186

 445 AGCATGTTCTCGCGCCGGGGAACTCTGCGGAGGCAAGACAAGAGACTCTGCAACGG  494
     ||||:::|||||||  :::::||||:::||||||||:::|||:::||||||
 187 SerMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAsnGl  203

 495 TGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGT  544
     ||||||||||||||||:::||  |||:::||||||||||||||||||||||
 203 yAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuVals  220

 545 CTTTCGGAAAAGCCCCGTGTGGCCAACTTGGCGTGCCAGGTGTCTACACC  594
     |:::::||  ||||||||||::|||||||||||::   |||::||||||||
 220 erTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr  236

 595 AACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCGTCCAG  636
     ||::::|||||||||||||||::::  |||:::  :::||:::
 237 GlnValCysLysPheThrLysTrpIleAsnAspThrMetLys  250
```

```
  68 lnAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGluAlaAspGln  84

 151 GAGCCAGGAGGCCAGATGGTGGAGGCCAGCCTCCGTACGGCACCCAGA  200
     ||||||||||||||:||||||||||||||:||||||||||||||
  85 GluProGlySerGlnMetValGluAlaSerLeuSerValArgHisProGl  101

 201 GTACAACAGACTCTTGCTCGCTAACGACCTCATCGCTCATCAAGTTGGACG  250
     ||||||||  :||||||||||||||||||:||||||||||:::|||
 101 uTyrAsnArgProLeuLeuAlaAsnAspLeuMet***IleLysLeuAspG  118

 251 AATCCGTGTCCGAGTCTGACACCATCCGGAGCATTGCTTCGCAG  300
     ||||||||||:::|||||||||||||:::||||||  |||:::||||
 118 luSerValSerGluSerAspAsnIleArg***IleSerIle***SerGln  134

 301 TGCCCTACCGCGGGGAACTCTTGCCTCGTNTCGGCTGGGGTCTCGCTGGC  350
     |||:::||||||||  |||||||  |||:||||:::||||||||||||||
 135 CysProThrAlaGlyAsnPheCysLeuValSerGlyTrpGlyLeuLeuAl  151

 351 GAACGGCAGAATGCCTACCGTCGCTGCCACTGCGTCGGTGGTGT  400
     |||||||||||||  |||||||||||:::|||||||||||||||
 151 aAsnGlyArgMetProThrValLeuGlnCysValAsnValSerValValS  168

 401 CTGAGGAGNTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATG  450
     ||||:  ||||||||||||||||||||||||||||||||||||||||
 168 erGluGluValCysSerLysLeuTyrAspProLeuTyrHisProSerMet  184

 451 TTCTGCGCCGCGGAGGGCCAAGACCAGAAGGACTCCTGCAACGGTGACTC  500
     ||||||||||||:::|||||||||||||||:::||||||||||||||||
 185 PheCysAlaGlyIleGlyAspSerGlnLysAspSerCysAsnGlyAspSe  201

 501 TGGGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTTCG  550
     |||||||||||||||||||||||||||||:::||||||||||||||||||
 201 rGlyGlyProLeuIleCysAsnGlyTyrLeuGlnGlyLeuValSerPheG  218

 551 GAAAAGCCCCGTGTGGCCAACTTGGCGTGCCAGTGTCTACACCAACCTC  600
     |||||||||:::|||||||||||||:::|||||||||||||||||||||
 218 lyLysAlaProCysGlyGlnValGlyValProGlyValTyrThrAsnLeu  234

 601 TGCAAATTCACTGAGTGGATAGAGAAAACCGTCCAGNCCAGT  642
     ||||||||||||||||:::  |||||||  ||||||||:::|||
 235 CysLysPheThrGluTrpIleGluLysThrValGlnAlaSer  248
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2

```
seq_documentation_block:
; Sequence 2, Application US/0855714.6
  Patent No. 5834290
; GENERAL INFORMATION:
;   APPLICANT:  Egelrud, Torbjorn
;   APPLICANT:  Hansson, Lennart
;   TITLE OF INVENTION:  Recombinant Stratum Corneum Chymotryptic
;   TITLE OF INVENTION:  Enzyme (SCCE)
;   NUMBER OF SEQUENCES:  17
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  White & Case, Patent Department
;     STREET:  1155 Avenue of the Americas
;     CITY:  New York
;     STATE:  New York
;     COUNTRY:  U.S.A.
;     ZIP:  10036-2787
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/557,146
;     FILING DATE:  14-DEC-1995
;     CLASSIFICATION:  424
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Sterner, Richard J.
;     REGISTRATION NUMBER:  35,372
;     REFERENCE/DOCKET NUMBER:  1103326-181
```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US96-04294-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
;   APPLICANT: Dixon, Eric P.
;   APPLICANT: Johnstone, Edward M.
;   APPLICANT: Little, Sheila P.
;   TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
;   TITLE OF INVENTION: RELATED NUCLEIC ACIDS
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Eli Lilly and Company
;   STREET: Lilly Corporate Center
;   CITY: Indianapolis
;   STATE: Indiana
;   COUNTRY: United States of America
;   ZIP: 46285
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US96/04294
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/416,257
;   FILING DATE: 04-APR-1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Blalock, Donna K.
;   REGISTRATION NUMBER: 38,082
;   REFERENCE/DOCKET NUMBER: X9239
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 317-276-1090
;   TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 253 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
PCT-US96-04294-2

alignment_scores:
        Quality: 530.50   Length: 214
          Ratio: 3.255   Gaps: 3
Percent Similarity: 76.168   Percent Identity: 47.196

alignment_block:
US-09-030-606-175 x PCT-US96-04294-2

Align seq 1/1 to: PCT-US96-04294-2   from: 1 to: 253

  1 GCGCAGCCCTGGCAGGCGGCACTGGTCATGCGAAAACGAATTGTTCTGCTC 50
    ::::::||||||||||||::::|||||| |||::||| |||::
 40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysGl 56

 51 GGGCGTCCTCGTGATCCGCAGTGGGTGCTGTCAGCCGCACACTGTTTCC 100
    :||||||||:: ::|||||||||::|||||||||||||
 56 yGlyValLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysLysM 73

101 AGAACTCCTACACATCGGGCTGGGCCTGCACAGTCTTGAGGCCGACCAA 150
    ::|||::|||||::: ||||| ||||| ||||:::
 73 etAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg 88

151 GAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCCTCCGTACGGCACCCAGA 200
    :: :::||| :::::||||||| ||| |||||||||
 89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProGl 103

201 GTACAACAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACG 250
    :::::: ||||||||||||::::||||||||||::|||||:::
103 yTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsns 120

251 AATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCCGCAG 300
    ::::::: ||:::::|| ::::::::::   :: ||:|||
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136

301 TGCCCTACCGCGGGAACTCTTGCCTCCGTNTCTGGCTGGGGTCTGCTGGC 350
    ||| ||| ||||::::||| :: |||||||||||
137 CysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153

351 GAACGGC......AGAATGCCTACCGTGCTGCGACTGCCTGAACGTGTCGG 394
    :::: :::|||:: ||| |||||::||:::::
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170

395 TGGTGTGTCGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCC 444
    ::::||| ||| :::||::|||:||| |||
170 euIleSerProGlnAspCysThrLysValTyrLysAspLeuLeuGluAsn 186

445 AGCATGTTCTGCGCCGGCGGAGGCAAGACCAGAAGGACTCCTGCAACGG 494
    |||||::|||||||| :::|||||||||| |||||||
187 SerMetLeuCysAlaGlyIleProAspSerLysAsnAlaCysAsnGl 203

495 TGACTCTGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGT 544
    |||||||||||||:::||::||| |||||||||||
203 yAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuValS 220

545 CTTTCGGAAAAGCCCGTGTGGCCAACTTGGCGTGCCAGGTGTCTACACC 594
    ::||| |||||||||||| ::: |||||||||||||
220 erTrpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr 236

595 AACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCGTCCAG 636
    ::::||||||||:|||||| ||||||||||
237 GlnValCysLysPheThrLysTrpIleAsnAspThrMetLys 250

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-557-146-12

seq_documentation_block:
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
;   APPLICANT: Egelrud, Torbjorn
;   APPLICANT: Hansson, Lennart
;   TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
;   TITLE OF INVENTION: Enzyme (SCCE)
;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: White & Case, Patent Department
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036-2787
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/557,146
;   FILING DATE: 14-DEC-1995
;   CLASSIFICATION: 424
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sterner, Richard J.
;   REGISTRATION NUMBER: 35,372
;   REFERENCE/DOCKET NUMBER: 1103326-181
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 819-8783
;   TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:

```
;     LENGTH:    225 amino acids
;     TYPE:      amino acids
;     STRANDEDNESS: single
;     TOPOLOGY:  linear
;     MOLECULE TYPE: polypeptide
US-08-557-146-12

alignment_scores:
          Quality:  516.00         Length:    215
          Ratio:    3.166          Gaps:      4
Percent Similarity:  75.814    Percent Identity: 46.512

alignment_block:
US-09-030-606-175 x US-08-557-146-12

Align seg 1/1  to: US-08-557-146-12   from: 1   to: 225

  1 GCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTG...TTCTG  47
    ::::::||||||||||::||||||:::    ||||:::||
 11 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisHisCy  27

 48 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT  97
    |:::||||||||||::::  ::||||||||:||||||||||||
 27 sGlyGlyValLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysL  44

 98 TCCAGAACTCCTACACCATCGGGCTGGGCTGCACAGTCTTGAGGCCGAC  147
    |||||||:||||::::|||||   :::|||     ::|||
 44 ysMetAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAsp  59

148 CAAGAGCCAGGAGCCAGGATGTGGAGGCCAGCCTCTCCGTACGGCACCC  197
    ::::::    ::||||   :::|||||  |||  ||||||
 60 ArgArg......AlaGlnArgIleLysAlaSerLysSerPheArgHisPr  74

198 AGAGTACAACAGACTCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGG  247
    |  |:|:     ::||||||||||||:::||:||
 74 oGlyTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuA  91

248 ACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCG  297
    ::::::::|||     ::||||   |||   :||||  ||||
 91 snSerGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSer 107

298 CAGTGCCCTACCGCGGGGAACTCTTGCCCGTNTCTGCGTGGGGTCTGCT  347
    ::::||   ::::::||||   ||||||||   |||||||||||
108 ArgCysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrTh 124

348 GGGCAACGGC......AGAATGCCTACCGTGCTGCACTGCGTGAACGTGT  391
    ::|||:::     ::||||::  :::||||     |||||||:::|||
124 rThrSerProAspValThrPheProSerAspLeuMetCysValAspVall 141

392 CGGTGGTCTGTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAC  441
    ::::::|||    |||||:::|||   ::|||     ::|
141 ysLeuIleSerProGlnAspCysThrGluValTyrLysAspLeuLeuGlu 157

442 CCCAGCATGTTCTGCGCCGGCGGAGGGCAAGACCAGAAGGACTCCTGCAA  491
    ||||||||||||    :::::||||||||    :::||||||||
158 AsnSerMetLeuCysSerAlaGlyIleProAspSerLysLysAsnAlaCysAs 174

492 CGTGACTCTGGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG  541
    ||||||||||||||||||||||||    |||||||:    ::||||||||
174 nGlyAspSerGlyGlyProLeuValCysArgGlyThrLeuGlnGlyLeuV 191

542 TGTCTTTCGGAAAAGCCCGGTGGCCAACTTGGCCCAGGTGTCTAC  591
    ||||::::|||     ||||||||     :::   ::||||:|||
191 alSerTrpGlyThrProPheCysGlyGlnProAsnAspProGlyValTyr 207

592 ACCAACCTCTGCAAATTCACTGATGGATAGAGAAAACCGTCCAG  636
    ||||:::|||||    :::|||     ||:::
208 ThrGlnValCysLysPheThrLysTrpIleAsnAspThrMetLys 222
```

```
;     Sequence 1, Application US/08361395
;     Patent No. 5733768
;     GENERAL INFORMATION:
;       APPLICANT:  Dixon, Eric P.
;       APPLICANT:  Johnstone, Edward M.
;       APPLICANT:  Little, Sheila P.
;       APPLICANT:  NO. 5733768ris, Franklin H.
;       TITLE OF INVENTION:  PROTEASE AND RELATED DNA COMPOUNDS
;       NUMBER OF SEQUENCES:  8
;       CORRESPONDENCE ADDRESS:
;         ADDRESSEE:  Eli Lilly and Company
;         STREET:  Lilly Corporate Center/Patent Division
;         CITY:  Indianapolis
;         STATE:  Indiana
;         COUNTRY:  United States
;         ZIP:  46285
;       COMPUTER READABLE FORM:
;         MEDIUM TYPE:  Floppy disk
;         COMPUTER:  IBM PC compatible
;         OPERATING SYSTEM:  PC-DOS/MS-DOS
;         SOFTWARE:  PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;         APPLICATION NUMBER:  US/08/361,395
;         FILING DATE:  22-DEC-1994
;         CLASSIFICATION:  530
;       ATTORNEY/AGENT INFORMATION:
;         NAME:  Blalock, Donna K.
;         REGISTRATION NUMBER:  38,082
;         REFERENCE/DOCKET NUMBER:  X8350A
;       TELECOMMUNICATION INFORMATION:
;         TELEPHONE:  317-277-1090
;         TELEFAX:  317-276-3861
;     INFORMATION FOR SEQ ID NO:    1:
;       SEQUENCE CHARACTERISTICS:
;         LENGTH:  244 amino acids
;         TYPE:  amino acid
;         STRANDEDNESS:  single
;         TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
US-08-361-395-1

alignment_scores:
          Quality:  470.50          Length:    215
          Ratio:    3.055           Gaps:      3
Percent Similarity:  71.628    Percent Identity: 42.326

alignment_block:
US-09-030-606-175 x US-08-361-395-1

Align seg 1/1  to: US-08-361-395-1  from: 1  to: 244

  1 GCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATGTTCTGCTC  50
    ::::::|||:::||||||||||     :::::::||||:::||:::
 32 SerHisProTyrGlnAlaAlaLeuTyrThrSerGlyHisLeuLeuCysGl  48

 51 GGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTCC 100
    :||||||||||:::||||  ||||||||||||||||||||||
 48 yGlyValLeuIleHisProLeuTrpValLeuThrAlaAlaHisCysLysL  65

101 AGAACTCCTACACACCATCGGGCTGGGCCTGCACAGTCTT.........GAG 141
    ::  :::   :::|||||||  |||||  :||||   ||||
 65 ysProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArgGlu  81

142 GCCGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACG 191
    ::::::|||||    ||||::::|||::::|
 82 SerSerGlnGlu...SerSerValArgAla........ValIll  94

192 GCACCCAGAGTACAACAGACTCTTGCTCGCTAACGACCTCATGCTCATCA 241
    :::||||||:||    ||||||:::   :::||:|||:::
 94 eHisProAspTyrAspAlaAlaSerHisAspGlnAspIleMetLeuLeuA 111

242 AGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATT 291
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-361-395-1

seq_documentation_block:

```
                              ; US-08-278-091-9
                                 TOPOLOGY: linear

                              alignment_scores:
                                        Quality:  456.50          Length:     211
                                          Ratio:    2.871            Gaps:       4
                              Percent Similarity:  75.355   Percent Identity: 41.706

                              alignment_block:
                              US-09-030-606-175 x US-08-278-091-9

                              Align seg 1/1  to: US-08-278-091-9   from: 1  to: 223 .

                                  7 CCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTCGGCGT   56
                                    |||::|||||::::|||          |||::::|||        ||||||::|||
                                 13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe   28

                                 57 CCTGGTGCATCCGACGTGGTGCTGCAGCCGCCACACTGTTCCAGAACT  106
                                    |||::::  |||||||::|||||||||::|||||||||||:::||||||
                                 28 rLeuIleAsnSerGlnTrpValValSerAlaAlaHisCysTyrLysSerG   45

                                107 CCTACACCATCGGGCGTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCA  156
                                    :::  |||||       :::::::::|||          |||
                                 45 lyIleGlnValArgLeuGlyGluAspAsnIleAsnVal...ValGluGly   60

                                157 GGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAA  206
                                    ::::::||||||||||  |||::||       ||||||:::|||
                                 61 AsnGluGlnPheIleSerAlaSerLysSerIleValHisProSerTyrAs   77

                                207 CAGACTCTTGCTCGCTAACGACCTCATCGTCATCAAGTTGGACGAATCCG  256
                                    |  |||||||||:::|||||||||||||||||||||||||||
                                 77 nSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuLysSerAlaA   94

                                257 TGTCCCAGTCTGCACCATCCGGAGCATCAGCATTGCTTCCCAGTGCCCT  306
                                    ::|||  :::::::  ::: |||||||:::   :::::::||
                                 94 laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla  110

                                307 ACCGCGGGAACTCTTGCCTCGTNTCTGGCTGGGTCTCTGCTGGCCAACGG  356
                                    ::||||||||:::|||||||:::|||||||||
                                111 SerAlaGlyThrGlnCysLeuIleSerGlyTyrpGlyAsnThrLysSerSe  127

                                357 C......AGAATGCCTACCGTGTGCCACTGCCGTGAACGTGTCGGTGGTGT  400
                                    ||      |||  |||||||||  |||||||||||:::::::  ||||||
                                127 rGlyThrSerTyrProAspValLeuLysCysLeuLysAlaProIleLeus  144

                                401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATG  450
                                    ||::::|| |||::::::  |||       |||       ::|||
                                144 erAspSerSerCysLysSerAlaTyrProGlyGlnIleThrSerAsnMet  160

                                451 TTCTGCGCCGGCGGAGGGCAAGACCAGAAGGACTCCTGCAACGGTGACTC  500
                                    ||||||||||      :::      ||||||||:::||||||||
                                161 PheCysAlaGlyTyrLeuGluGlyLysAspSerCysGlnGlyAspSe  177

                                501 TGGGGGCCCCTGATCTGCAACGGTACTTGCAGGGCCTTGTGTCTTTCG  550
                                    |||||||||:::||||||:::|||  :::||||||  ||||||||||||
                                177 rGlyGlyProValValCysSerGlyLysLeuGlnGlyIleValSerTrpIg  194

                                551 GAAAAGCCCCGTGTGCCAACTTGGCGTGCCAGGTGTCTACCACCCTC  600
                                    |||::::::  ||::::|||  :::  ||||||||||||||:::::||
                                194 lySerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal  209

                                601 TGCAAATTCACTGAGTGGATAGAGAAAAACCGTC  633
                                    |||||||||:::|||||||::::||:::||
                                210 CysAsnTyrValSerTrpIleLysGlnThrIle  220

                              seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-483-859-9

                              seq_documentation_block:
                              ; Sequence 9, Application US/08483859
                              ; Patent No. 5656436
                              ; GENERAL INFORMATION:
                              ;     APPLICANT: LOOSMORE, Sheena M.
```

```
111 rgLeuAlaArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu 127
    ::|||     :::   :::::::    |||:::      :::       :::
292 GCTTCGCAGTGCCTACCGCGGGGAACTCTTGCCTCGTNTCTGGCTGGGG 341
    ::::|||    :::        :::|||||||  :::||||||
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTrpGl 144
    |   ||:::|||    :::|||    :::::::|||:::
342 TCTGCTGGCGAACGGCAGAATGCCTACCGTGCTGCACTCCGTGAACGTGT 391
    ||yLysThrAlaAspGlyAspPheProAspThrIleGlnCysAlaTyrIleH 161
392 CGGTGGTGTCTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCAC 441
    :::::|||:::      |||:::      |||
161 isLeuValSerArgGluGluCysGluHisAlaTyrProGlyGlnIleThr 177
442 CCCAGCATGTTCTGCGCCGGCGGAGGGCAAGACCAGAAGGACTCCTGCAA 491
    :::||||:::||||||||      :::        ||||||||::|||
178 GlnAsnMetLeuCysAlaGlyTyrLeuGlyLysTyrGlyLysAspSerCysGl 194
492 CGGTGACTCTGGGGGCCCCTGATCTGCAACGGTACTTGCAGGGCCTTG 541
    :::||||||||||||||||:::|||:::       :::|||:::||||||
194 nGlyAspSerGlyGlyProLeuValCysGlyAspHisLeuArgGlyLeuV 211
542 TGTCTTTCGGAAAAGCCCCGTGTGCCAACTTGGCGTGCCAGGTGTCTAC 591
    |||||:::|||       |||:::|||:::
211 alSerTrpIleGlyAsnIleProCysGlySerLysGluLysProGlyValTyr 227
592 ACCAACCCTGCAAATTCACTGAGTGGATAGAGAAAAACCGTCCAG 636
    ||||:::::::||::::|||||||||:::||||||||||||  :::|||
228 ThrAsnValCysArgTyrThrAsnTrpIleGlnLysThrIleGln 242
```

```
seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-278-091-9

seq_documentation_block:
; Sequence 9, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
;     APPLICANT: LOOSMORE, Sheena M
;     APPLICANT: YANG, Yan-Ping
;     APPLICANT: CHONG, Pele
;     APPLICANT: OOMEN, Raymond P.
;     APPLICANT: KLEIN, Michel H.
;     TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
;     TITLE OF INVENTION: Reduced Protease Activity
;     NUMBER OF SEQUENCES: 23
;     CORRESPONDENCE ADDRESS:
;         ADDRESSEE:  Sim & McBurney
;         STREET:  Suite 701, 330 University Avenue
;         CITY:  Toronto
;         STATE:  Ontario
;         COUNTRY:  Canada
;         ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;         MEDIUM TYPE:  Floppy disk
;         COMPUTER:  IBM PC compatible
;         OPERATING SYSTEM:  PC-DOS/MS-DOS
;         SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;         APPLICATION NUMBER:  US/08/278,091
;         FILING DATE:  21-JUL-1994
;         CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;         NAME:  Stewart, Michael I
;         REGISTRATION NUMBER:  24,973
;         REFERENCE/DOCKET NUMBER:  1038-371
;     TELECOMMUNICATION INFORMATION:
;         TELEPHONE:  (416) 595-1155
;         TELEFAX:  (416) 595-1163
; INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH:  223 amino acids
;         TYPE:  amino acid
;         STRANDEDNESS:  single
```

```
;   APPLICANT:    YANG, Yan-Ping
;   APPLICANT:    CHONG, Pele
;   APPLICANT:    OOMEN, Raymond P.
;   APPLICANT:    KLEIN, Michel H.
;   TITLE OF INVENTION:  Analog of Haemophilus Hin47 Protein with
;   TITLE OF INVENTION:  Reduced Protease Activity
;   NUMBER OF SEQUENCES:  23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Sim & McBurney
;     STREET:  Suite 701, 330 University Avenue
;     CITY:  Toronto
;     STATE:  Ontario
;     COUNTRY:  Canada
;     ZIP:  M5G 1R7
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0,  Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/483,859
;     FILING DATE:  07-JUN-1995
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/296,149
;     FILING DATE:  26-AUG-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/278,091
;     FILING DATE:  21-JUL-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Stewart, Michael I.
;     REGISTRATION NUMBER:  24,973
;     REFERENCE/DOCKET NUMBER:  1038-495 MIS:vg
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (416) 595-1155
;     TELEFAX:  (416) 595-1163
; INFORMATION FOR SEQ ID NO:  9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  223 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
US-08-483-859-9

alignment_scores:
      Quality:  456.50             Length:  211
      Ratio:  2.871                Gaps:  4
Percent Similarity:  75.355    Percent Identity:  41.706

alignment_block:
US-09-030-606-175 x US-08-483-859-9

Align seg 1/1  to: US-08-483-859-9  from: 1  to: 223

    7 CCCTGCCAGGCGGCCACTGGTCATGCAAACGAATTGTTCTGCTCGGGCGT 56
      |||:|||:||||:::|||      :::::    ||||||::||
   13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe 28

   57 CCTGGTGCATCCGCAGTGGGTGCTGTGCAGCCGCCACACTGTTTCCAGAACT 106
      |||:::::   |||||||||::||||||||||||:::::::::
   28 rLeuIleAsnSerGlnTrpValValSerAlaAlaHisCysTyrLysSerG 45

  107 CCTACACCATCGGGCTGGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCA 156
      :: ||||||   |:|||||    :::::::||   |:::::: |||
   45 lyIleGlnValArgLeuGlyLeuGlyGluAspAsnVal...ValGluGly 60

  157 GGGAGCCAGATGGTGGAGGCCAGCCTTCCGTACGGCACCCAGAGTACAA 206
      :::::||||:::||||||::|||||   |||||::||||||:::||||
   61 AsnGluGlnPheIleSerAlaSerLysSerIleValHisProSerTyrAs 77

  207 CAGACTCTTGCTCGCTAACGACCTCATCGCTCATCCAAGTTGGACGAATCCG 256
      |::       ||| ||||||:::|||||||| |||||||||||
```

```
   77 nSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuLysSerAlaA 94
      ::|||   :::::   :|   |||||:::   ||    |::|||||:::||
  257 TGTCCGAGTCGACACCATCCGGACATCAGCATTGCTTCGCAGTGCCCT 306
      ::|||   :::::   :|   |||||||||::    :::|||||:::||
   94 laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla 110

  307 ACCGGCGGGAACTCTTGCCTCGTCNTCTGCTGGGTCTGCTGGCGAACGG 356
      ::|||||||:::||||:::|||||::||||||||||||     :::::
  111 SerAlaGlyThrGlnCysLeuIleSerGlyTyrProGlyAsnThrLysSerSe 127

  357 C......AGAATGCCTACCGTGCTGCACTGCTGCTGAACGTGTCGGTGGTGT 400
      :||    :||:::::   |||:::::::   :::::|
  127 rGlyThrSerTyrProAspValLeuLysCysLeuLysAlaProIleLeuS 144

  401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATG 450
      ||:::::|   |||||::|||:::   |||
  144 erAspSerCysLysSerAlaTyrProGlyGlnIleThrSerAsnMet 160

  451 TTCTGCCCGGGCGGAGGGCAAGACCAGAAGGACTCCTGCAACGGTGACTC 500
      ||||||||||     :::   |||||||||||:::||||||:::||||||
  161 PheCysAlaGlyTyrLeuGluGlyGluGlyLysAspSerCysGlnGlyAspSe 177

  501 TGGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTCG 550
      |||||||||:::||||||||||||::|||||:::|||||||:::||||
  177 rGlyGlyProValValCysSerGlyLysLeuGlnGlyIleValSerTrpG 194

  551 GAAAAGCCCCGTGTGGCCAACTTGGCGTGCCCAGGTGTCTACCACCAACCTC 600
      ||:::::|   :::||||   :::   |||||||||:::|||||||||:::
  194 lySerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal 209

  601 TGCAAATTCACTGAGTGGATAGAGAAAACCGTC 633
      ||:::::|   :::|||||::|||||:::|||
  210 CysAsnTyrValSerTrpIleLysGlnThrIle 220
```

```
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9

seq_documentation_block:
;  Sequence 9, Application US/08472173
;  Patent No. 5665353
;  GENERAL INFORMATION:
;    APPLICANT:    LOOSMORE, Sheena M
;    APPLICANT:    YANG, Yan-Ping
;    APPLICANT:    CHONG, Pele
;    APPLICANT:    OOMEN, Raymond P.
;    APPLICANT:    KLEIN, Michel H.
;    TITLE OF INVENTION:  Analog of Haemophilus Hin47 Protein with
;    TITLE OF INVENTION:  Reduced Protease Activity
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sim & McBurney
;      STREET:  Suite 701, 330 University Avenue
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5G 1R7
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0,  Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,173
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,149
;      FILING DATE:  26-AUG-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/278,091
;      FILING DATE:  21-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stewart, Michael I.
;      REGISTRATION NUMBER:  24,973
```

Wed Sep 29 14:27:27 1999

```
;     REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  (416) 595-1155
;        TELEFAX:    (416) 595-1163
;  INFORMATION FOR SEQ ID NO: 9:
;     SEQUENCE CHARACTERISTICS:
;        TYPE: amino acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
US-08-472-173-9

alignment_scores:
        Quality:   456.50          Length:      211
        Ratio:     2.871           Gaps:          4
Percent Similarity:  75.355    Percent Identity: 41.706

alignment_block:
US-09-030-606-175 x US-08-472-173-9    ..

Align seg 1/1  to: US-08-472-173-9  from: 1  to: 223

  7 CCCTGGCAGGCGGCACTGGTCATGAGAAAACGAATTGTTCTGCTCGGGCGT  56
    |||::||:::|||       :::::|   |||||||:::||||
 13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe  28

 57 CCTGGTGCATCCGAGTGGGTGCTGTCAGCCGCACACTGTTTCCAGAACT   106
    |||:||||::  |||||||||:||||||||||||||||::|||||::||
 28 rLeuIleAsnSerGlnTrpValSerAlaAlaHisCysTyrLysSerG     45

107 CCTACCACCATCGGCTGGGCCTGCACAGTCTTGAGGCGCGACCAAGAGCCA  156
    ::  ||||| |||||       ::::::::       ::           ||
 45 lyIleGlnValArgLeuGlyGluAspAsnIleAsnVal...ValGluGly  60

157 GGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAA  206
    :::||||:||:::|||||:::|||   ||||:::   ::|||||
 61 AsnGluGlnPheIleSerAlaSerLysSerIleValHisProSerTyrAs  77

207 CAGACTCTTGCTGTCTAACGACCTCATGCTCATCAAGTTGGACGAATCCG  256
    |    ||| |||||::||||:||||||||||||||||   :::::::|||
 77 nSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuLysSerAlaA  94

257 TGTCCGAGTCTGACACCATCCGAGACATCAGCATTGCTTCCCAGTGCCCT  306
    ::||| ||::   ::  ||||||||||   |||:::     ::|||||||
 94 laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla  110

307 ACCGCGGGGAACTCTTGCCTCGTNTCTGGCTGGGGTCTCGCTGGCGAACGG  356
    :::||||||||   :: ||||||||||       ::::::|||
111 SerAlaGlyThrGlnCysLeuIleSerGlyTrpGlyAsnThrLysSerSe  127

357 C......AGAATGCCTACCGTGCCCACTGCGTGAACGTGTCGGTGGTGT   400
    ||   |||||    |||||     :::|||||||||:::|||:::|
127 rGlyThrSerTyrProAspValLeuLysCysLeuLysAlaProIleLeus  144

401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCGTACCACCCAGCATG  450
    ||:::::   |||:||||    |||           ::|||
144 erAspSerCysLysSerAlaTyrProGlnIleThrSerAsnMet    160

451 TTCTGCGCCGCGGAGGCAAGACCAGAAGGACTCCTGCAACGGTGACTC  500
    ||||||||||   ::   ||||||||||||||:::|||||||
161 PheCysAlaGlyTyrLeuGluGlyGlyLysAspSerCysGlnGlyAspSe  177

501 TGGGGGCCCCTGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTTCG  550
    ::|||||||::||:::||||||||||||||||||||||:::
177 rGlyGlyProValCysSerGlyLysLeuGlnGlyIleValSerTrpG    194

551 GAAAAGCCCCGTGTGGCCAACTGGCGCTGCCAGGTGTCTCACCAACCTC  600
    ||::::::   |||:|| || ||||||||||||||||||||
194 lySerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal  209

601 TGCAAATTCACTGAGTGGATGGAATAGAGAAAACCGTC  633
    ||:::::::::||||||:::::||||
```

---

210 CysAsnTyrValSerTrpIleLysGlnThrIle 220

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-487-167-9

```
seq_documentation_block:
; Sequence 9, Application US/0848716709
; Patent No. 5869302
; GENERAL INFORMATION:
;     APPLICANT: LOOSMORE, Sheena M.
;     APPLICANT: YANG, Yan-Ping
;     APPLICANT: CHONG, Pele
;     APPLICANT: OOMEN, Raymond P.
;     APPLICANT: KLEIN, Michel H.
;     TITLE OF INVENTION:  Analog of Haemophilus Hin47 Protein with
;     TITLE OF INVENTION:  Reduced Protease Activity
;     NUMBER OF SEQUENCES:  23
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE:  Sim & McBurney
;        STREET:  Suite 701, 330 University Avenue
;        CITY:  Toronto
;        STATE:  Ontario
;        COUNTRY:  Canada
;        ZIP:  M5G 1R7
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC compatible
;        OPERATING SYSTEM:  PC-DOS/MS-DOS
;        SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/08/487,167
;        FILING DATE:  07-JUN-1995
;        CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: US 08/296,149
;        FILING DATE:  26-AUG-1994
;     PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: US 08/278,091
;        FILING DATE:  21-JUL-1994
;     ATTORNEY/AGENT INFORMATION:
;        NAME:  Stewart, Michael I.
;        REGISTRATION NUMBER:  24,973
;        REFERENCE/DOCKET NUMBER:  1038-508 MIS:vg
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  (416) 595-1155
;        TELEFAX:   (416) 595-1163
;  INFORMATION FOR SEQ ID NO: 9:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH:  223 amino acids
;        TYPE:  amino acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
US-08-487-167-9

alignment_scores:
        Quality:   456.50          Length:      211
        Ratio:     2.871           Gaps:          4
Percent Similarity:  75.355    Percent Identity: 41.706

alignment_block:
US-09-030-606-175 x US-08-487-167-9  from: 1  to: 223  ..

Align seg 1/1  to: US-08-487-167-9  from: 1  to: 223

  7 CCCTGGCAGGCGGCACTGGTCATGAGAAAACGAATTGTTCTGCTCGGGCGT  56
    |||::||:::|||       ::::::   ||||||||||:::|||
 13 ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe  28

 57 CCTGGTGCATCCGAGTGGGTGCTGTCAGCCGCACACTGTTTCCAGAACT  106
    |||:||||::  ||||||||::||||||||||||||||::|||||::|||
 28 rLeuIleAsnSerGlnTrpValSerAlaAlaHisCysTyrHisPheGlySe  45

107 CCTACCACCATCGGCTGGGCCTGCACAGTCTTGAGGCGCGACCAAGAGCCA  156
    ::  ||||| |||||       ::::::::       ::           ||
```

ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 871814
US-08-744-026-3

alignment_scores:
; Quality: 456.00   Length: 224
; Ratio: 2.980   Gaps: 3
; Percent Similarity: 68.304   Percent Identity: 40.179

alignment_block:
US-09-030-606-175 x US-08-744-026-3

Align seg 1/1 to: US-08-744-026-3   from: 1 to: 258

```
  1 GGCCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTC 50
    ::|||||||||||||||||||||        :: ::  |||||     |||
 32 SerGlnProTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCysGl 48

 51 GGGGGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACTGTTTCC 100
    :|||::||||:||::||||:||||||||::|||||||||:||||||||::
 48 yGlyIleLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleG 65

101 AGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGACCAA 150
    ::::::::||     |||||| ||||| |||||          ||:::||
 65 lyAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAspAsp 80

151 GAGCCAGGGAGCCAGAATGGTGGAGGCCAGCCTCCCGTACGGCACCCAGA 200
    |||        :::||||:::|||||:||||:||| |||    ||||||||
 81 GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProCy 97

201 GTACAACAGACTCTTGCTCGCTAAC......................... 225
    ::||| |||| ||||                               :: ||
 97 sPheAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAspGluAspT 114

226 ........GACCTCATGCTCATCAGTTGAGCAGAATCCGTGTCCGAGTCT 267
         ||||||||||||||||       :: |||:::: |||||||||
114 yrSerHisAspSerHisLeuMetLeuLeuArgLeuThrGlnProAlaGlu 130

268 GACACCATCCGGAGCATCAGCAATTGCTTCCCAGTGCCCTACCGCGGGAA 317
    |||:::|||||     |||||:::|||||   :::||| |||  ::|||::
131 AspAlaValGlnValGluGluProThrGlnGluProGluValGlySe 147

318 CTCTTGCCTCCGTNTCTGGCTGGGGTCTGCTG......GCGAACGGCAGAA 361
    :::::|||||:|||||||||||||||||   ::     ||| |||
147 rThrCysLeuAlaSerGlyThrTrpGlySerIleGluProGluAsnPheSerT 164

362 TGCCTACCGTGCTGCACTGCGTGAACGTGCGGTGGTGTCTGAGGANGTC 411
    |||   |||||||::|||||:::::::::::     ::
164 yrProAspAspSerGlnCysValAspSerLeuLysIleLeuProAsnAspLys 180

412 TGCAGTAAGCTCTATGACCCCGCTGTACCACCCCAGCATGTTCTGCGCCGG 461
    ||:::|||  :::::||       :::||| ||||||||:::|||||||||
181 CysAlaLysAlaHisThrGlnLysValThrGluPheMetLeuCysAlaGl 197

462 CGGAGGGCAAGAACCAGAAGGAGCTCCTGCAACGTGACTCTGGGGGCCCC 511
    ::: |||||||||||||||||||||||:::||| |||||||||||||||||
197 yHisLeuGlyGlyLysAspThrCysValGlyAspSerGlyGlyProL 214
```

seq_documentation_block:
; Sequence 3, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-744-026-3

```
 45 lylleGlnValArgLeuGlyGluAspAsnIleAsnVal...ValGluGlyGly 60
    ::    |||||               :::::::::::           |||
157 GGGAGGCCAGATGGTGGAGGCCAGCCCTCCGTACGGCACCCAGAGTACAA 206
    ::::::||:::|||:::||||| |||||  ||||:  |||||:: |||||
 61 AsnGluGlnPheIleSerAlaSerLysSerIleValHisProSerTyrAs 77

207 CAGACTCTTGCTCGCTAACGACCTCATGCTCATCAGTTGACGAGAATCCG 256
    |||  |||   ||||||||||||||||||||||||||||| :::::::::
 77 nSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuLysSerAlaA 94

257 TGTCCAGTGTCTGACACCATCCGGAGCATCAGCAATTGCTTCCAGTGCCCT 306
    :::||  ::::::   ::  ||||||||||:::  :::::::|||
 94 laSerLeuAsnSerArgValAlaSerIleSerLeuProThrSerCysAla 110

307 ACCGCGGGAACTCTTGCCTCGTNTCTGGCTGGGGTCTGCTGGCCGAACGG 356
    :::||||||:::::||||:::||||||||||||||||||||||     :::
111 SerAlaGlyThrGlnCysLeuIleSerGlyThrTrpGlyAsnThrLysSerSe 127

357 C......AGAATGCCTACCGTGCTGCACTGCGTGAACGTGTCGGTGGTGT 400
        ||||||  ||||| ||||||||||:::    :::::::|:|
127 rGlyThrSerTyrProAspValLeuCysLeuLysAlaProIleLeus 144

401 CTGAGGANGTCTGCAGTAAGCTCTATGACCCCGCTGTACCACCCCAGCATG 450
    |||:::||   ::||||:||    ::::||:::::::::      ::|||
144 erAspSerCysLysSerAlaTyrProGlyGlnIleThrSerAsnMet 160

451 TTCTGCGCCGGCGAGGGCAAGAACCAGAAGGACTCCTGCAACGGTGACTC 500
    |||||||||||            ::     ||||||||||||:::||||||
161 PheCysAlaGlyTyrLeuGluGlyLyGlyLysAspSerCysGlnGlyAspSe 177

501 TGGGGGCCCCTGATCTGCAACGGTACTTGCGAGGGCCTTGTGTCTTTCG 550
    |||||||||::|||  ||||||||:||   |||||||::|||:::|||:::|
177 rGlyGlyProValValCysSerGlyLeuGlnGlyIleValSerTrpG 194

551 GAAAGCCCCGTGTGGCCAACTTGGCGTGCGGTGCTGTCTACACCAACCTC 600
    ||::::::  |||:   ||:::|||  |||||::::||:||||||||||
194 lySerGly...CysAlaGlnLysAsnLysProGlyValTyrThrLysVal 209

601 TGCAAATTCACTGAGTGGATAGAGAAAACCGTC 633
    |||:||||:::|||||:::::||||:||:::
210 CysAsnTyrValSerTrpIleLysGlnThrIle 220
```

```
512 TGATCTGCAACGGGTACTTGCAGGGCCTTGTGTCTTCGGAAAAGCCCCG 561
    || :::||||||| ||||||| |||||||::::||::||
214 euThrCysAspGlyValLeuGlnGlyValThrSerTrpGlyTyrIlePro 230

562 TGTGGCCAACTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAATTCAC 611
    ||||::: ::: ||::||::||||||||:::||||
231 CysGlySerProAsnLysProAlaValPheValArgValLeuSerTyrVa 247

612 TGAGTGGATAGAGAAAACCGTC 633
    ::||||||:: ||||:||
247 lLysTrpIleGluAspThrIle 254
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-11

seq_documentation_block:
; Sequence 11, Application US/08096946
; Patent No. 5516639
; GENERAL INFORMATION:
;   APPLICANT: Tindall, Donald J.
;   APPLICANT: Young, Charles Y-F
;   APPLICANT: Klee, George G.
;   TITLE OF INVENTION: Antibodies Specific for Human
;   TITLE OF INVENTION: Prostate Glandular Kallikrein
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Merchant & Gould
;   STREET: 3100 No. 5516639west Center
;   CITY: Minneapolis
;   STATE: MN
;   COUNTRY: USA
;   ZIP: 55402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/096,946
;   FILING DATE: 19930722
;   CLASSIFICATION: 436
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Woessner, Warren D.
;   REGISTRATION NUMBER: 30,440
;   REFERENCE/DOCKET NUMBER: 1016.62-US-01
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 612-332-5300
;   TELEFAX: 612-332-8081
; INFORMATION FOR SEQ ID NO: 11:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 237 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   ORIGINAL SOURCE:
;   ORGANISM: Amino acid sequence of PSA (hK3)
US-08-096-946-11

alignment_scores:
       Quality: 450.00   Length: 224
         Ratio: 2.922   Gaps: 3
Percent Similarity: 68.750   Percent Identity: 39.286

alignment_block:
US-09-030-606-175 x US-08-096-946-11

Align seg 1/1 to: US-08-096-946-11 from: 1 to: 237

  1 GCGGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTC 50
    :::||||||||| ||| ::: :::||::: |||::
 11 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysGl 27
```

```
 51 GGGCGTCCTGGTGCATCCGAGTGGTGGCTCGTCAGCCGCCACACTGTTTCC 100
    :::|||||||||||||||||||||||||||| :::||||||||||||||||
 27 yGlyValLeuValHisProGlnTrpLeuThrAlaAlaHisCysIleA 44

101 AGAACTCCTACACCATCGGGCTGGGCCTGCAGAGTCTTGAGCGCGACCAA 150
    ::||::: ::: ||||| |||||| |||||||
 44 rgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisProGlu 60

151 GAGCCAGGGAGGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA 200
    :::: ||| ||::: :::||::||:::|| |||||
 61 AspThrGly...GlnValPheLeuValSerThrSerPheProHisProLe 76

201 GTACAACAGACTCTTGCTGCTAAC...................... 225
    ||::: ||| |||
 76 uTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAspS 93

226 ........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCT 267
    ||||||||::::|||   :::::: :::
 93 erSerHisAspLeuMetLeuArgLeuSerGluProAlaGluLeuThr 109

268 GACACCATCCGGAGCATCAGCATTGCTTCCAGTGCCCTACCGCCGGGGAA 317
    ||||||||||||||::: ::: ||||||| |||::: |||
110 AspAlaValLysValMetAspLeuProThrGlnProAlaLeuGlyTh 126

318 CTCTTGCCTCGTNTCTGGCTGGGGTCTGCTGGCGAACGGCAGAATG.... 363
    :::||: :::||||||| ::: ::: ::::::
126 rThrCysTyrAlaSerGlyTrpGlyLysSerIleGluProGluGluPheLeuT 143

364 ..CCTACCGTGCTGCGCCTGCGAACGTGTCGGTGGTGTCTGAGGANGTC 411
    ||| ::::||||||||::: ||
143 hrProLysLysGlnCysValGlnLeuHisLeuHisSerAsnAspVal 159

412 TGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATGTTCTGCCGCCGG 461
    :::::::::::: ||| :::||||||||
160 CysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAlaGl 176

462 CGGAGGCAAGACCAGAAGGACTCCTGCAACGGTGACTCTGGGGGCCCCC 511
    | :::|||::::|||||||||||||||||||
176 yArgTrpThrGlyGlyLysSerGlyAspSerGlyGlyProL 193

512 TGATCTGCAACGGGTACTTGCAGGGGCCTTGTGTCTTCGGAAAAGCCCCG 561
    ||::::|||||||| :::|||||| ::: :::||::||
193 euValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlySerGluPro 209

562 TGTGGCCAACTTGGCGTGCTCTACACCAACCTCTGCAAATTCAC 611
    |||::: ::: ||||||||||||||
210 CysAlaLeuProGluArgProSerLeuTyrThrLysValHisTyrAr 226

612 TGAGTGGATAGAGAAAACCGTC 633
    ::||||||:: ||||:||
226 gLysTrpIleGluAspThrIle 233
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11

seq_documentation_block:
; Sequence 11, Application PC/TUS9407329
; GENERAL INFORMATION:
;   APPLICANT: Mayo Foundation for Medical
;   APPLICANT: Education and Research
;   TITLE OF INVENTION: Antibodies Specific for Human
;   TITLE OF INVENTION: Prostate Glandular Kallikrein
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Schwegman, Lundberg & Woessner
;   STREET: 3500 IDS Center
;   STREET: 80 South Eighth Street
;   CITY: Minneapolis
;   STATE: MN
;   COUNTRY: USA
;   ZIP: 55402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible

```
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: PCT/US94/07329
      FILING DATE:
      CLASSIFICATION:
    ATTORNEY/AGENT INFORMATION:
      NAME: Raasch, Kevin W.
      REGISTRATION NUMBER: 35,651
      REFERENCE/DOCKET NUMBER: 150.62WO
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 612-339-0331
      TELEFAX: 612-339-3061
  INFORMATION FOR SEQ ID NO: 11:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 237 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: peptide
    ORIGINAL SOURCE:
      ORGANISM: Amino acid sequence of PSA (hK3)
PCT-US94-07329-11

alignment_scores:
       Quality:    450.00          Length:      224
       Ratio:        2.922         Gaps:          3
Percent Similarity:  68.750   Percent Identity: 39.286

alignment_block:
US-09-030-606-175 x PCT-US94-07329-11

Align seg 1/1  to: PCT-US94-07329-11  from: 1  to: 237

  1 GCGCGCCCCTGGCCAGGCGGCGCACTCGTCATGGAAAAACGAATTGTTCTGCTC 50
    :::|||||||||||::   :::::  ::::::::  ||::::::
 11 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysGl 27

 51 GGGCGTCCTCGTGCATCCCAGTGGTGCTGTCAGCCCGACCACTGTTTCC 100
    :||||||||||||||||||::|||||||||||||||||::
 27 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleA 44

101 AGAACTCCTACACACCGGCTGGCCTGCACAGTCTTGAGGCCGACCAA 150
    :::|||::::  :::|||  |||||||||||||::::||||
 44 rgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisProGlu 60

151 GAGCCAGGGAGGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA 200
    :::   |||  |||:::    :::::::||| |||     |||||
 61 AspThrGly...GlnValPheGlnValSerThrSerPheProHisProLe 76

201 GTACAACAAGACTCTTGCTCGCTAAC...................... 225
    ||::: |||||  |||||
 76 uTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAspS 93

226 ........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCT 267
    |||||||||||  :::::::||| :::::::  |||
 93 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThr 109

268 GACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAA 317
    ||:::::::  ::::::::::  :::::  |||  ||| |||:::
110 AspAlaValLysValMetAspLeuProThrGlnCysProAlaLeuGlyTh 126

318 CTCTTGCCTCGTNTCTGGCTGGGGTCTGCTGGGCGAACGGCAGAATG.... 363
    :::|||||||||   ::  :::::
126 rThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheLeuT 143

364 ..CCTACCGTCGCCACTGCCGTGAACGTGTCGGTGGTGTCTGAGGANGTC 411
    |||   |||:::  ||::|||||||:::  :||:::  |||
143 hrProLysLeuGlnCysValGlnLeuHisValIleSerAsnAspVal 159

412 TGCAGTAAGCTCTATGACCCGCTGTACCCCCCAGCATGTTCTGCGCCGG 461
    |+|::::::::::   |||::::||||||||
```

```
160 CysAlaGlnValHisProGlnLysLysValThrLysPheMetLeuCysAlaGl 176

462 CGGAGGGCAAGACCAGAAGGACTCCTGCAACGGTGACTCTGGGGGGCCCC 511
    |||::::::||||::::::|||||||||||||||||||||||
176 yArgTrpThrGlyGlyLysSerThrCysSerGlyAspSerGlyGlyProL 193

512 TGATCTGCAACGGTACTTGCAGGGCCTTGTGTCTTTCGGAAAAGCCCCG 561
    ||::||||||||   |||||||||   :::|||||::::||| |||
193 euValCysAsnGlyValLeuGlnGlyLyIleThrSerTrpGlySerGluPro 209

562 TGTGGCCAACTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAATTCAC 611
    |||:::  :::::::::||||||::::::||||||||   :::
210 CysAlaLeuProGluArgProSerLeuTyrThrLysValHisTyrAr 226

612 TGAGTGGATAGAGAAAACCGTC 633
    :::|||||:::  |||:::
226 glySerTrpIleLysAspThrIle 233
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-1

```
alignment_scores:
       Quality:    450.00          Length:      224
       Ratio:        2.922         Gaps:          3
Percent Similarity:  68.750   Percent Identity: 39.286

alignment_block:
US-09-030-606-175 x PCT-US95-06157-1

Align seg 1/1  to: PCT-US95-06157-1  from: 1  to: 237
```

```
   1 GCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTC  50
     :::|||||||||||||::||  :||:::::||  |||||
  11 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysGl  27

  51 GGGCGTCCTGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACTGTTCC  100
     ::||||||||||||||||||||::||:::||::||:::|||:::
  27 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleA  44

 101 AGAACTCCTACACCACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGACCAA  150
     ::|||::::  ::|||  |||||  ||||||   :::
  44 rgAsnLysSerValIleLeuLeuGlyLeuGlyArgHisSerLeuProHisProGlu  60

 151 GAGCCAGGAGGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA  200
     :::  ||| :::||  |||:::::|||  ||||||
  61 AspThrGly...GlnValPheGlnValSerThrSerProPheProHisProLe  76

 201 GTACAACAGACTCTTGCTCGCTAAC....................  225
     |||::|
  76 uTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAsps  93

 226 .......GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCT  267
     |||||||||||||||::|||   :::
  93 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThr  109

 268 GACACCATCCGGAGCCATCAGCATTGCTTCCGCAGTGCCCTACCGCGGGGAA  317
     |||::::||  ::::::||   :|||  |||:||  |||
 110 AspAlaValLysValMetAspLeuProThrGlnLeuProAlaLeuGlyTh  126

 318 CTCTTGCCTCGTNTCTGGGGTTCGGCTCTGGTGGTGTCTGAGGANGTC.....  363
     ::|||  :::|||||||  |:::::|:::||:::|
 126 rThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheLeuT  143

 364 ..CCTACCGTGCTGCACTGCCTGAACGTGTCCGGTGGTGTCTGAGGANGTC  411
       |||   |||:::|||||:::::::  |||:::|||:: :::
 143 hrProLysLysGluGlnCysValGlnLeuHisValIleSerAsnAspVal  159

 412 TGCAGTAAGCTCTATGACCCCCTGTACCACCCCAGCATGTTCTGCGCCGG  461
     |||::|||||||  ||||||||||||:::::||::|||:::||||||
 160 CysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAlaGl  176

 462 CGGAGGGCAAGACCAGAAGGACTCCTGCAACGGTGACTCCTGGGGGGCCCC  511
     ||||::||||||||||||||||||||:::||::||||||||||||
 176 yArgTrpThrGlyLyLysSerThrCysSerGlyAspSerGlyGlyProL  193

 512 TGAATCTGCAACGGGTACTTGCCAGGGCCTTGTGTCTTCGGAAAAGCCCCG  561
     |::||||||  |||||||||||||||:::||:::|||  :::
 193 euValCysAsnGlyValLeuGlnIleThrSerTrpSerGluPro  209

 562 TGTGGCCAACTTGGCGTGCCAGGTGTCTACACCAACCTCTGCAAATTCAC  611
     :::::||||||||||::||||||:::
 210 CysAlaLeuProGluArgProSerLeuTyrThrLysValHisTyrAr  226

 612 TGAGTGGATAGAGAAAACCGTC  633
     ::|||||||||||  |||:::
 226 gLysTrpIleLysAspThrIle  233
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-744-026-5

```
alignment_scores:
                 Quality:   449.00           Length:    224
                   Ratio:    2.916             Gaps:      3
   Percent Similarity:  68.750   Percent Identity:  39.286

alignment_block:
US-09-030-606-175 x US-08-744-026-5

Align seg 1/1  to: US-08-744-026-5  from: 1  to: 261

   1 GCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTC  50
     :::|||||||||||||:::  :::::::  |||
  35 SerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCysGl  51

  51 GGGCGTCCTGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTCC  100
     ::||||||||||||||||||||||||||||||||:::||::
  51 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleA  68

 101 AGAACTCCTACACCACCATCGGGCTGGGCCTGGGCCTGCACAGTCTTGAGGCCGACCAA  150
     ::|||||  |||||:::  |||||  |||||   :::
  68 rgAsnLysSerValIleLeuLeuGlyLeuGlyArgHisSerLeuProHisProGlu  84

 151 GAGCCAGGAGGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA  200
     :::  |||  :::|||  |||:::::|||  ||| |||||
  85 AspThrGly...GlnValPheGlnValSerHisSerProPheProHisProLe  100

 201 GTACAACAGACTCTTGCTCGCTAAC..................  225
     |||::  |||
 100 uTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspAsps  117

 226 .......GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCT  267
     ||||||:::||:::|||   :::
 117 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeuThr  133

 268 GACACCATCCGGAGCCATCAGCATTGCTTCCGCAGTGCCCTACCGCGGGGAA  317
     |||::|  :::||||  ::::||   :::
 134 AspAlaValLysValMetAspLeuProThrGlnLeuProAlaLeuGlyTh  150

 318 CTCTTGCCTCGTNTCTGGGGTTCGGCTCTGGTGGTGTCTGCGCGAACGGGCAGAATG.....  363
     ::|||  :::|||||||  :::::|::||
```

Wed Sep 29 14:27:27 1999

us-09-030-606-175.rai

Percent Similarity: 67.411   Percent Identity: 67.411   Ratio: 2.967   Gaps: 3   39.286

alignment_block:
US-09-030-606-175 x PCT-US95-06157-8

Align seg 1/1 to: PCT-US95-06157-8   from: 1 to: 238

```
    1 GCGCAGCCCTGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTGCTC   50
      ::::||||||||||||||:::::    :::::   |||||||:::
   12 SerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCysGl   28

   51 GGGCGGTCCTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTTTCC  100
      :||||||||||||||||||||||||||||||||||||||||||||||
   28 yGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysLeuL   45

  101 AGAACTCCTACACCATCGGGCTGGGCCTCCACAGTCTTGAGGCGACCAA  150
      :::::::   :::  |||||| :::::||
   45 ySLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluProGlu   61

  151 GAGCCAGGGAGCCCAGATGGTGGAGCCAGCCTCCCGTACGGCACCCAGA  200
      ::: |||  :::||   :::|||   |||||
   62 AspThrGly...GlnArgValProValSerHisSerPheProHisProLe   77

  201 GTACAACAGACTCTTGCTC.......................   219
      ||||||
   77 uTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluAspS   94

  220 ..GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGGTCCGAGTCT  267
      :: ||||||||||||||::||||:::|||
   94 erSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIleThr  110

  268 GACACCATCCGGAGCACAGCATTGCTTGCAGTGCCCTACCGCGGGGAA  317
      |||:::::::   :::::||||||:::   :::||  |||:::|||||
  111 AspValLysValLeuGlyLeuProThrGlnGluProAlaLeuGlyTh  127

  318 CTCTTGCCTCGTNTCTGGCTGGGGTCTGCTCGGCGAACGGCAGAATG.....  363
      :: ::|||||||||||    :::  ::::::
  127 rThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheLeuA  144

  364 ..CCTACCGTCGCCGCTGCCACTGCGTGAACGTGTCCGGTGGTGTCTGAGGANGTC  411
      |||  ||||||||:::::|||  :::|||
  144 rgProArgSerLeuGlnCysValSerLeuHisLeuSerAsnAspMet  160

  412 TGCAGTAAGCTCTATGACCCGCTGTACCACCCAGCATGTTCTGCGCCGG  461
      |||:::|||:: |||::  |||||||::||||||
  161 CysAlaAArgAlaTyrSerGluLysValThrGluPheMetLeuCysAlaGl  177

  462 CGGAGGGCAAGACCAGAAGGACTCCTGCAACGTGACTCTGGGGGGCCCC  511
      |||||||||||||:::|||||||||:::||||||||||||||||
  177 yLeuTrpThrGlyGlyLysAspThrCysGlyLyGlyAspSerGlyGlyProL  194

  512 TGATCTGCAACGGGTACTTGCAGGGCCTTGTGCTTTCGGAAAAGCCCCG  561
      !!::|||||||    ||||||||::||||||||:::||||:::|||
  194 euValCysAsnGlyValLeuGlnGlyIleThrSerTrpGlyProGluPro  210

  562 TGTGGCCAACTTGGCGTGCCAGTGTCTACACCAACCTCTGCAAATTCAC  611
      ||:::|||||||||   |||::
  211 CysAlaLeuProGluLysProAlaValTyrThrLysValHisTyrAr  227

  612 TGAGTGGATAGAGAAAACCGTC  633
      ::||||||:::  |||:::
  227 gLysTrpIleLysTyrThrIle  234
```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US95-06157-8

alignment_scores:
; Quality: 448.00   Length: 224

; Quality: 448.00

us-09-030-606-175.rai

This Page Blank (uspto)

OM of: US-09-030-606-177 to: Issued_Patents_AA:*   out_format : pfs

Date: Sep 25, 1999 4:01 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/USPTO_spool/US09030606/runat.24091999_171616_29813/app_query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=score
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-177
Query length: 1119
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

score_list:

```
Sequence                                                      Strd  Orig   ZScore  EScore   Len | Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1              +    610.00 1152.47 1.7e-57  248 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2             +    274.50  505.33 1.9e-21  253 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2        +    274.50  505.33 1.9e-21  253 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-1            +    272.50  501.06 3.1e-21  262 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-3            +    272.50  501.06 3.1e-21  262 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-681-951-1           +    272.50  501.06 3.1e-21  262 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-11          +    272.00  501.29 3.3e-21  237 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11       +    272.00  501.29 3.3e-21  237
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-795-06157-1         +    272.00  501.29 3.3e-21  237
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-1        +    270.50  497.39 5.0e-21  258 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-3           +    270.50  496.28 5.7e-21  261 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-744-026-5           +    268.50  493.35 8.3e-21  262 !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-4           +    266.50  490.64 1.3e-20  238 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8        +    266.50  490.34 1.3e-20  244 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-10       +    266.50  489.54 1.4e-20  261 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-6        +    264.50  486.83 2.1e-20  237
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-096-946-10          +    264.50  486.83 2.1e-20  237
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-10       +    264.50  486.83 2.1e-20  237
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-16       +    264.50  486.68 2.1e-20  240 !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-228A-1          +    264.50  486.68 2.1e-20  237 !
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-09303-1        +    264.00  485.87 2.4e-20  237
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-844-024-2           +    260.50  479.74 5.5e-20  225
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-12          +    250.00  459.33 7.5e-19  228
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-278-091-7           +    250.00  459.33 7.5e-19  228
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-7           +    250.00  459.33 7.5e-19  228
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-7           +    250.00  459.33 7.5e-19  228
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-487-167-7           +    244.00  448.03 3.3e-18  223
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-9           +    244.00  448.03 3.3e-18  223
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-483-859-9           +    244.00  448.03 3.3e-18  223
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-472-173-9           +    244.00  448.03 3.3e-18  223
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-487-167-9           +    231.00  419.99 9.3e-17  286
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-9          +    231.00  419.99 9.3e-17  286
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-9          +    231.00  419.99 9.3e-17  286
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-9        +    229.50  418.10 1.3e-16  263
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-4          +    224.00  405.97 5.4e-16  299
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-8          +    224.00  405.97 5.4e-16  299
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-8          +    224.00  405.97 5.4e-16  299
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-278-091-8          +    221.00  403.21 9.8e-16  232
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-990-301A-4          +    219.50  399.72 1.5e-15  244
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-361-395-1          +    219.00  397.07 1.8e-15  281
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-1          +    219.00  397.07 1.8e-15  281
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-7          +    219.00  397.00 1.8e-15  281
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-7        +    218.00  395.57 2.2e-15  271
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-467-155A-10          +    218.00  395.57 2.2e-15  271
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-628-198-10          +    218.00  395.57 2.2e-15  271
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-07343-10       +    218.00  395.57 2.2e-15  271
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-1

seq_documentation_block:
; Sequence 1, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
;   APPLICANT: Bandman, Olga
;   APPLICANT: Goli, Surya K.
;   TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
;   TITLE OF INVENTION: KALLIKREIN
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: US
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/744,026
;   FILING DATE: Herewith
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0154 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 248 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   LIBRARY:
;   CLONE: Consensus
US-08-744-026-1

alignment_scores:
;   Quality: 610.00   Length: 149
;   Ratio: 4.692   Gaps: 2
;   Percent Similarity: 87.248   Percent Identity: 80.537

alignment_block:
US-09-030-606-177 x US-08-744-026-1

Align seg 1/1 to: US-08-744-026-1 from: 1 to: 248

```
    4 CACTCGGCAGCCCTGGCAGGCGGCACTGGTCATGGAACACGAATTGTTCTG  53
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   34 HisSerGlnProTrpGlnAlaAlaLeuValMetGluAsnGluLeuPheCy   50

   54 CTTCGGGCGTCCTGGTCATCCCGCAGTGGTGCTTGTCAGCCGCCACACTGTT 103
      |||||||||||||||||||||||||||||||||||||||||||||||||
   50 sSerGlyValLeuValHisProGlnTrpValLeuSerAlaAlaHisCysP   67

  104 TCCAGAACTCCTACACCATCGGCGTGGGCCTGGCCGTCTTGAGGCCGAC  153
      |||||||||||||||||||||||||||||||||||||||||||||||||
```

67 heGlnAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGluAlaAsp 83

154 CAAGAGCCAGGAGCCAGATGTGGAGGCCAGCCTCCGTACGGCCACCC 203
    ||||||||||||||||||||||||||||||||||||||||||||||
84 GlnGluProGlySerGlnMetValGluAlaSerLeuSerValArgHisPr 100

204 AGAGTACAACAGACCCTGCTCGCTAACGACCTCATGCTCATCAAGTTGG 253
    |||||||||||||||||||||||||||||||:||||||| |||||||||
100 OGluTyrAsnArgProLeuLeuAlaAsnAspLeuMet***IleLysLeuA 117

254 ACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCG 303
    |||||||||||||||||||||||::|||||||||||:::|||||
117 spGluSerValSerGluSerAspAsnIleArg***IleSerIle***Ser 133

304 CAGTGCCCTACCGCGGGGGAACTCTTGCCTCGTTTCTGCGGTGGGGTCTGCT 353
    ||||||||||||||||||||||||||||||||||||||||
134 GlnCysProThrAlaGlyGlyAsnPheCysLeuValSerGlyTrpGlyLeuLe 150

354 GGCGAACGATGCTGTG.............................ATTGCCATCC 379
    |||||| :::
150 uAlaAsnGlyArgMetProThrValLeuGlnCysValAsnValSerValV 167

380 AGTCCCAGACTGTGGGAGGCTGGGAGTGTGAGAAGCTTTCCCAACCC 426
    |||::: ||| |||::||||| :::||
167 alSerGluGluVal...........CysSerLysLeuTyrAspPro 178

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-557-146-2

alignment_scores:
        Quality: 274.50   Length: 143
        Ratio: 2.745   Gaps: 3
Percent-Simrilarity: 69.930   Percent Identity: 38.462

---

alignment_block:
US-09-030-606-177 x US-08-557-146-2

Align seg 1/1 to: US-08-557-146-2 from: 1 to: 253

7 TCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTCTGCTC 56
  |||::||||||||||||:::|||||::: |||::||| |||||
40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysGl 56

57 GGGCGTCCTGGTGCATCCGCAGTGGTGCTGTCAGCCCACACTGTTTCC 106
   ||||||||||||||:::   ::||||||||||||||||:||||||
56 yGlyValLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysLysM 73

107 AGAACTCCTACACCATCGGGCTGCGGCTGCCACAGTCGTTGAGGCCGACCAA 156
    ::||||::||||||::||| ||||| ::||||::|||
73 etAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg 88

157 GAGCCAGGGAGCCAGGATGGTGGAGGCCAGCCTCCGTACGGCACCCAGA 206
    :::   ::||| :::||||:|||||||| ||||||||
89 Arg.....AlaGlnArgLeuLysAlaSerLysSerPheArgHisProGl 103

207 GTACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGACG 256
    ||||::   ::|||||||||||||||||||||::|||:
103 yTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsns 120

257 AATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCGCAG 306
    :::::::: |||::: :::::::::::::::|::   ::| |||:::
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136

307 TGCCCTACCGCGGGGAACTCTTGCCTCGTTTCTGCGGTGGGGTCTGCTGGC 356
    ||| :::::||||:::||| ||||||||||| ::
137 CysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh 153

357 GAACGATGCTGTG.......................ATTGCCATCCAGTCCCAGA 388
    ::: ||| ::| ::
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170

389 CTGTGGGAGGCTCGGAGTGTGAGAAGCTT 417
    :::::: :::||| |||:::
170 euIleSerProGlnAspCysThrLysVal 179

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US96-04294-2

```
;    NAME: Blalock, Donna K.
;    REGISTRATION NUMBER: 38,082
;    REFERENCE/DOCKET NUMBER: X9239
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 317-277-1090
;      TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 253 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
PCT-US96-04294-2

alignment_scores:
        Quality:    274.50               Length:      143
        Ratio:        2.745              Gaps:          3
Percent Similarity:  69.930     Percent Identity: 38.462

alignment_block:
US-09-030-606-177  to  PCT-US96-04294-2

Align seg 1/1  to:  PCT-US96-04294-2   from:  1  to: 253     .

    7  TCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTGCTC   56
       |||||||||||||||||||||||||:::|||         |||:::|||  |||
   40  SerHisProTrpGlnValAlaLeuSerGlyAsnGlnLeuHisCysGl   56

   57  GGGCGTCCTGGTCGCATCCGCAGGTGGTGCTGTCAGCCGCCACACTGTTCC  106
       :|||||||||||:  ::|||||||||||||||:|||||||||||||
   56  yGlyValLeuValAsnGluArgTrpValLeuThrAlaAlaHisCysLysM   73

  107  AGAACTCCTACACCATCGGGCTGGGCCTCGCACAGTCTTGAGGCCGACCAA  156
       :::||::|||||::||||  ||||||    ::|||    ::||::::
   73  etAsnGluTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg   88

  157  GAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGA  206
       ::|       ||||||       :::::::::    :::  ||||||||
   89  Arg.....AlaGlnArgIleLysAlaSerLysPheArgHisProGl  103

  207  GTACAACAGACCCTTGCTCGCTAACGACCTCATGCTCATCAAGTTGGACG  256
       :: ||::   |||||||||||||  ||||||||||:::|||||||||||
  103  yTyrSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsnS  120

  257  AATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAGCATTGCTTCGCAG  306
       :::::::::   |||:::|  :::::::::|||   :::|||||
  120  erGlnAlaArgLeuSerSerMetValLysValArgLeuProSerArg  136

  307  TGCCCTACCGCGGGGAACTCTTCGCCTCGTTTCTGGCTGGGGTCTGCTGGC  356
       |||::|||||:::||  |||||||||  :::||
  137  CysGluProProGlyThrThrCysThrValSerGlyTrpGlyThrThrTh  153

  357  GAACGATGCTGTG...........ATTGCCATCCAGTCCCCAGA  388
       ::|    ||||||             :::||||||::::  ::::
  153  rSerProAspValThrProSerAspLeuMetCysValAspValLysL  170

  389  CTGTGGGAGGCTGGGAGTGTGAGAAGCTT  417
       ::|||      ::|||   |||:::
  170  euIleSerProGlnAspCysValThrLysVal  179

seq_name:  /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-790-137-1

seq_documentation_block:
; Sequence 1. Application US/08790137
; Patent No. 5840871
; GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goll, Surya K.
;    TITLE OF INVENTION:  A NOVEL PROSTATE-ASSOCIATED
;    TITLE OF INVENTION:  KALLIKREIN
;    NUMBER OF SEQUENCES:  4
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Incyte Pharmaceuticals, Inc.
;      STREET: 3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY: USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Diskette
;      COMPUTER: IBM Compatible
;      OPERATING SYSTEM: DOS
;      SOFTWARE: FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/790,137
;      FILING DATE: Filed Herewith
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Billings, Lucy J.
;      REGISTRATION NUMBER: 36,749
;      REFERENCE/DOCKET NUMBER: PF-0195 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 415-855-0555
;      TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  262 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
US-08-790-137-1

alignment_scores:
        Quality:    272.50               Length:      155
        Ratio:        2.698              Gaps:          4
Percent Similarity:  65.161     Percent Identity: 38.710

alignment_block:
US-09-030-606-177  to  US-08-790-137-1

Align seg 1/1  to:  US-08-790-137-1   from:  1  to: 262     .

    4  CACTCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTG   53
       |||||||||||||||||||||||||||        ::  ||  ||
   34  HisSerGlnProTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCy   50

   54  CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT  103
       |:::||:::|||||||         |||||||||:::||||||||||||
   50  sGlyGlyIleLeuValHisArgGlnTrpValLeuThrAlaAlaHisCysI   67

  104  TCCAGAACTCCTACACCATCGGGCTGGGCCTGCCACAGTCTTGAGGCCGAC  153
       :::::::|||    ||||:::|||    |||:||    ||::::|||
   67  leSerAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAsp   82

  154  CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC  203
       :::|||       :::|||:::::||||:::|||    |||  ||||||
   83  AspGluAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr   99

  204  AGAGTACACAGACACCCTTGCTCGCTAAC  231
       :: ||||   ||||||||||||  |||
   99  oGlyPheAsnMetSerLeuLeuGluAsnHisThrArgGlnAlaAspGluA  116

  232  ..........GACTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG  270
                 ||||||||||||:::|||   :::::
  116  spTyrSerHisAspLeuMetLeuLeuArgLeuThrGluProAlaAspThr  132

  271  ...TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTCCCTACCGC  317
       :::|||||:||||:::|||   :::|||||
  133  IleThrAspAlaValLysValValGluLeuProThrGlnGluProGluVa  149
```

Wed Sep 29 14:27:33 1999

```
 34  HisSerGlnProTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCy   50
 54  CTCGGGCGTCCTGGTCGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTT  103
     |::||::||::|||||||||  |||||||||||:|||::||||||||||||
 50  sGlyGlyIleLeuValHisArgGlnTrpValLeuThrAlaAlaHisCysI   67

104  TCCAGAACTCCTACCACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC  153
     ::::::::::||  :::  |||||| |||:::||| |||:::|||  |||
 67  leSerAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAsp   82

154  CAAGAGGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC  203
     ::|||  ||||  ||||||||||||| ||||| |||||||||  |||||
 83  AspGluAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr   99

204  AGAGTACAACAGACCCTTGCTCGCTAAC.................  231
     :::|||  |||||| |||
 99  oGlyPheAsnMetSerLeuLeuGluAsnHisThrArgGlnAlaAspGluA  116

232  .........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG  270
              ||||||||||||||:::||| |||  :::::
116  spTyrSerHisAspLeuMetLeuLeuArgLeuThrGluProAlaAspThr  132

271  ...TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGC  317
     ::::||||||:||  :::::::  :::::::|||  ||||||||  ::
133  IleThrAspAlaValLysValValGluLeuProThrGlnGluProGluVa  149

318  GGGGAACTCTTCGCCTCGTTTCTGCGGGGTCGCTGGCG..........  357
     :||::||::|||||:::::||||||||||||
149  lGlySerThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnP  166

358  ........AACGATGCTGTGATTGCCATCCAGTCCCAGACTGTGGGAGGC  399
             ||  :::  ::::::::| |||  ||| :::
166  heSerPheProAspProAspLeuGlnCysValAspLeuLysIleLeuProAsn  182

400  TGGGAGTGTGAGAAG  414
     ||||||||||||
183  AspGluCysGluLys  187
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-681-151-4

```
; seq_documentation_block:
; Sequence 4, Application US/08681151
; Patent No. 5869637
; GENERAL INFORMATION:
;   APPLICANT:  Au-Young, Janice
;   APPLICANT:  Bandman, Olga
;   APPLICANT:  Braxton, Scott Michael
;   APPLICANT:  Goli, Surya
;   TITLE OF INVENTION:  A NOVEL HUMAN KALLIKREIN
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;     STREET:  3174 Porter Drive
;     CITY:  Palo Alto
;     STATE:  CA
;     COUNTRY:  US
;     ZIP:  94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/681,151
;     FILING DATE:  Herewith
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Billings, Lucy J.
;     REGISTRATION NUMBER:  36,749
;     REFERENCE/DOCKET NUMBER:  PF-0074US
```

```
318  GGGGAACTCTTCGCCTCGTTTCTGCGGGGTCGCTGGCG..........  357
     :||::||::|||||:::::||||||||||||
149  lGlySerThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnP  166

358  ........AACGATGCTGTGATTGCCATCCAGTCCCAGACTGTGGGAGGC  399
             ||  :::  ::::::::| |||  ::: :::
166  heSerPheProAspProAspLeuGlnCysValAspLeuLysIleLeuProAsn  182

400  TGGGAGTGTGAGAAG  414
     ||||||||||||
183  AspGluCysGluLys  187
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-790-137-3

```
; seq_documentation_block:
; Sequence 3, Application US/08790137
; Patent No. 5840871
; GENERAL INFORMATION:
;   APPLICANT:  Hillman, Jennifer L.
;   APPLICANT:  Goli, Surya K.
;   TITLE OF INVENTION:  A NOVEL PROSTATE-ASSOCIATED
;   TITLE OF INVENTION:  KALLIKREIN
;   NUMBER OF SEQUENCES:  4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;     STREET:  3174 Porter Drive
;     CITY:  Palo Alto
;     STATE:  CA
;     COUNTRY:  USA
;     ZIP:  94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/790,137
;     FILING DATE:  Filed Herewith
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Billings, Lucy J.
;     REGISTRATION NUMBER:  36,749
;     REFERENCE/DOCKET NUMBER:  PF-0195 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415-855-0555
;     TELEFAX:  415-845-4166
; INFORMATION FOR SEQ ID NO:  3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  262 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   IMMEDIATE SOURCE:
;     LIBRARY:  GenBank
;     CLONE:  186653
; US-08-790-137-3
```

```
alignment_scores:
     Quality:  272.50          Length:  155
       Ratio:  2.698            Gaps:  4
Percent Similarity:  65.161    Percent Identity:  38.710

alignment_block:
US-09-030-606-177 x US-08-790-137-3    .
Align seg 1/1  to: US-08-790-137-3  from: 1 to: 262

   4  CACTCGCAGCCCTGCCAGGCGGCACTGGTCATGGAAACAGAATTGTTCTG  53
      |||:||||||||||||||||||||||||||||||||||||||||||||||  ||
```

```
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415-855-0555
;     TELEFAX:   415-845-4166
;     TELEX:
;   INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  262 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   HYPOTHETICAL:  NO
;   IMMEDIATE SOURCE:
;     LIBRARY:  GENBANK
;     CLONE:  125170
US-08-681-151-4
```

```
alignment_scores:
        Quality:  272.50              Length:       155
          Ratio:   2.698                Gaps:         4
Percent Similarity:  65.161    Percent Identity:  38.710
```

```
alignment_block:
US-09-030-606-177 x US-08-681-151-4    .
```

```
Align seg 1/1  to:  US-08-681-151-4   from:  1  to: 262

   4 CACTCGCAGCCCTGGCAGGCCGGCACTGGTCATGGAAAACGAATTGTTCTG  53
     ||||||||||||||||||||||||||||        :::  ::::    ||
  34 HisSerGlnProTrpGlnAlaAlaLeuTy:HisPheSerThrPheGlnCy  50

  54 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 103
     |:::|||::|||||||:::|||||||||||||::::||||:||:|||||||
  50 sGlyGlyIleLeuValHisArgGlnTrpValLeuThrAlaAlaHisCysI  67

 104 TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC 153
     ::::::::||||  ::::  ||||||  ||||| |||:::|||  :||||
  67 leSerAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAsp  82

 154 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 203
     ::::||        :::||||::||::::||| |||  |||   ||||
  83 AspGluAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr  99

 204 AGAGTACACACAGACCCTTGCTCGCTAAC................... 231
     |  ::|||:||||   ||||| |||
  99 oGlyPheAsnMetSerLeuLeuGluAsnHisThrArgGlnAlaAspGluA 116

 232 .........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
     |||||||||||::::|||  |:::||
 116 spTyrSerHisAspLeuMetLeuLeuArgLeuThrGluProAlaAspThr 132

 271 ...TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGC 317
     :::|||::::::  :::::::|   :::||| ||| :::
 133 IleThrAspAlaValLysValGlyValGluLeuProThrGlnProGluVa 149

 318 GGGAACTCTTGCCTCGTTTCTGGCTGGGGTCTGCTGGCG.......... 357
     |  :::||||::||:|||||| ||||| |:::
 149 lGlySerThrCysLeuAlaSerGlyTyrTrpGlySerIleGluProGluAsnP 166

 358 .......AACGATGCTGATTGCCATCCAGTCCAGGTGTGGGAGGC 399
     |||||||||||:::    :::  ::  |||| |||
 166 heSerPheProAspLeuGlnCysValAspLeuLysIleLeuProAsn 182

 400 TGGGAGTGTGAGAAG 414
     |||||||||||
 183 AspGluCysGluLys 187
```

```
seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-095946
```

```
seq_documentation_block:
; Sequence 11, Application US/08095946
; Patent No. 5516639
```

```
;   GENERAL INFORMATION:
;     APPLICANT:  Tindall, Donald J.
;     APPLICANT:  Young, Charles Y-F
;     APPLICANT:  Klee, George G.
;   TITLE OF INVENTION:  Antibodies Specific for Human
;   TITLE OF INVENTION:  Prostate Glandular Kallikrein
;   NUMBER OF SEQUENCES:  11
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Merchant & Gould
;     STREET:  3100 No. 5516639west Center
;     CITY:  Minneapolis
;     STATE:  MN
;     COUNTRY:  USA
;     ZIP:  55402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0,  Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/096,946
;     FILING DATE:  19930722
;     CLASSIFICATION:  436
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Woessner, Warren D.
;     REGISTRATION NUMBER:  30,440
;     REFERENCE/DOCKET NUMBER:  1016.62-US-01
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  612-332-5300
;     TELEFAX:   612-332-8081
;   INFORMATION FOR SEQ ID NO:  11:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  237 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   ORIGINAL SOURCE:
;     ORGANISM:  Amino acid sequence of PSA (hK3)
US-08-096-946-11
```

```
alignment_scores:
        Quality:  272.00              Length:       170
          Ratio:   2.720                Gaps:         4
Percent Similarity:  58.824    Percent Identity:  34.706
```

```
alignment_block:
US-09-030-606-177 x US-08-096-946-11    ..
```

```
Align seg 1/1  to:  US-08-096-946-11   from:  1  to: 237

   4 CACTCGCAGCCCTGGCAGGCCGCACTGGTCATGGAAAACGAATTGTTCTG  53
     |||||||||||||||||||||||||||||  :::::  :::::::    |
  10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyAlaValAlaValCy  26

  54 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 103
     |:::|||::|||||||:::|||||||||||::::||||:||:|||||||||
  26 sGlyGlyValHisValHisProGlnTrpValLeuThrAlaAlaHisCysI  43

 104 TCCAGAACTCCTACACCATCGGGCTGGGCCTGGCCTGCACAGTCTTGAGGCCGAC 153
     :::::||:::||::  :::  ||||||  ||||||  |||:::|||  |||||
  43 leArgAsnLysSerValIleLeuValGlyArgHisSerLeuPheHisPro  59

 154 CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC 203
     ::::||    :::  :::||||::||::::|||  |||   ||| ::|
  60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr  75

 204 AGAGTACACAGACCCTTGCTCGCTAAC.................... 231
     |  |||::    ||||| |||
  75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA  92

 232 .........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
```

alignment_block:
US-09-030-606-177 x PCT-US94-07329-11

Align seg 1/1 to: PCT-US94-07329-11 from: 1 to: 237

```
  4 CACTCGCAGCGCCTGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG 53
    |||||||||||||||||||::::  .:::::: ::::::::
 10 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 26

 54 CTCGGCGTCCTGCTGCATCCGCAGTGGGTGTCAGCCGCACACTGTT 103
    |:::|||||||||||||||::::::::::: ||||||||||
 26 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 43

104 TCCAGAACTCCTACACCATGCGGCCTGGGCTGCCCACAGTCTTGAGGCCGAC 153
    :::::||||:: ::||| ||||| |||||||
 43 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 59

154 CAAGAGCCAGGAGGCCAGATGGGAGGCCAGCCTCTCCGTACGGCACCC 203
    ::::: ||| |||::.:::::||| || |||||
 60 GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr 75

204 AGAGTACAACAGACCCTTGCTGCTAAC... 231
    | |||::: ||||| ||| |
 75 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 92

232 .........GACCTCATGCTCATCAAGTTGGACGAACCGTGTCCGAG 270
    ||||||||||||||||::::|||
 92 spSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108

271 TCTGACACCATCCGGAGCCATCAGCATTGCTTCGCAGTGCCTACCGCGGG 320
    ::|||||||||||||:::   :::|||  |||::: |
109 ThrAspAlaValLysMetAspLeuProThrGlnGluProAlaLeuGl 125

321 GAACTCTTGCCTCGTTTCTGGCTGGGGT... 348
    |:::::||| :|||||||||||
125 yThrThrCysTyrAlaSerGlyTrpSerIleGluProGluGluPheL 142

349 ...... CTGCTGGCGAACGAT 363
    :::::::::::||||||
142 euThrProLysLeuGlnCysValGlnLeuHisValIleSerAsnAsp 158

364 GCTGTGATTGCCATCCAGTCCCAGACTGTG......... 393
    ::: |||:::||||||||||||
159 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 175

394 .GGAGGCTGG 402
    ||| |||
175 aGlyArgTrp 178
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-1

```
 92 spSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 108
    ||||||||||||::::|||||:::|||
271 TCTGACACCATCCGGAGCCATCAGCATTGCTTCGCAGTGCCTACCGCGGG 320
    |:::::||| :||||||||||
109 ThrAspAlaValLysMetAspLeuProThrGlnGluProAlaLeuGl 125

125 yThrThrCysTyrAlaSerGlyTrpSerIleGluProGluGluPheL 142
    ::::::||| :||| |||:::: |
321 GAACTCTTGCCTCGTTTCTGGCTGGGGT 348

349 ...... CTGCTGGCGAACGAT 363

142 euThrProLysLeuGlnCysValGlnLeuHisValIleSerAsnAsp 158
    ::::::::::::|||||
364 GCTGTGATTGCCATCCAGTCCCAGACTGTG 393

159 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 175
    :::::::||||||
394 .GGAGGCTGG 402

175 aGlyArgTrp 178
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US94-07329-11

alignment_scores:
```
    Quality:  272.00           Length:      170
      Ratio:    2.720             Gaps:        4
Percent Similarity: 58.824   Percent Identity: 34.706
```

```
;      OPERATING SYSTEM:   PC-DOS/MS-DOS
;      SOFTWARE:   PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:   PCT/US95/06157
;      FILING DATE:
;    CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:   Raasch, Kevin W.
;      REGISTRATION NUMBER:   35,561
;      REFERENCE/DOCKET NUMBER:   150.148WO1
;    TELECOMUNICATION INFORMATION:
;      TELEPHONE:   612-339-0331
;      TELEFAX:   612-339-3061
;  INFORMATION FOR SEQ ID NO:   1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:   237 amino acids
;      TYPE:   amino acid
;      STRANDEDNESS:   single
;      TOPOLOGY:   linear
;    MOLECULE TYPE:   peptide
PCT-US95-06157-1

alignment_scores:
         Quality:   272.00              Length:   170
         Ratio:   2.720                 Gaps:   4
Percent Similarity:   58.824   Percent Identity:   34.706

alignment_block:
US-09-030-606-177 x PCT-US95-06157-1

Align seg 1/1 to: PCT-US95-06157-1   from: 1   to: 237

     4  CACTCGCAGCCCTGCCAGGCGGCCACTGGTCATGCAGGAAAACGAATTGTTCTG   53
        |||||||||||||||::   :::::   :::::::||||   |||
    10  HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy   26

    54  CTCGGGGGTCCTGGTCGCATCCCAGTGGTGCTGTCAGCCGCACACTGTT   103
        |::::|||||||||||||||||||||||::||||||||||||||||
    26  sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI   43

   104  TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACAGTCTTGAGGCCGAC   153
        :::||||||   :::||   |||||||||||   |||||
    43  leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro   59

   154  CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC   203
        :::::   |||   |||::   ||||||::||||   |||   |||||
    60  GluAspThrGly...GlnValPheGlnValSerThrSerPheProHisPr   75

   204  AGAGTACAACACGACCCTTGCTCGCTAAC..............   231
        |   ||::   |||||   ||
    75  OLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA   92

   232  ..........GACCTCATGCTCATCAAGTGGACGAATCCGTGTCCGAG   270
                  ||||||||||||||:::::::||::|||
    92  spSerSerHisAspLeuMetLeuArgLeuSerGluProAlaGluLeu   108

   271  TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGG   320
        ::||||||||   :::||||   ||::   :||   |||
   109  ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGl   125

   321  GAACTCCTTGCCTCGTTTCTGGCTGGGT................   348
        |   ||:::|||   ::|||||||||||
   125  yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL   142

   349  ..........................CTGCTGGCGAACCGAT   363
                                  :::::|||||||
   142  euThrProLysLysGlnCysValGlnLeuHisLeuHisValIleSerArgAsnAsp   158

   364  GCTGTGATTGCCATCCAGTCCCAGACTGTG................   393
        :::::   |||   |||||||||||
   159  ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl   175
```

```
;   394 .GGAGGCTGG 402
;          ||| ||| |||
;   175 aGlyArgTrp 178
;
; seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-744-026-3
;
seq_documentation_block:
; Sequence 3, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
;      APPLICANT:   Bandman, Olga
;      APPLICANT:   Goli, Surya K.
;      TITLE OF INVENTION:   A NOVEL PROSTATE-SPECIFIC
;      TITLE OF INVENTION:   KALLIKREIN
;      NUMBER OF SEQUENCES:   5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:   INCYTE PHARMACEUTICALS, INC.
;      STREET:   3174 Porter Drive
;      CITY:   Palo Alto
;      STATE:   CA
;      COUNTRY:   US
;      ZIP:   94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:   Diskette
;      COMPUTER:   IBM Compatible
;      OPERATING SYSTEM:   DOS
;      SOFTWARE:   FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:   US/08/744,026
;      FILING DATE:   Herewith
;      CLASSIFICATION:   514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:   Billings, Lucy J
;      REGISTRATION NUMBER:   36,749
;      REFERENCE/DOCKET NUMBER:   PF-0154 US
;    TELECOMUNICATION INFORMATION:
;      TELEPHONE:   415-855-0555
;      TELEFAX:   415-845-4166
; INFORMATION FOR SEQ ID NO:   3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:   258 amino acids
;      TYPE:   amino acid
;      STRANDEDNESS:   single
;      TOPOLOGY:   linear
;    IMMEDIATE SOURCE:
;      LIBRARY:   GenBank
;      CLONE:   871814
US-08-744-026-3

alignment_scores:
         Quality:   270.50              Length:   159
         Ratio:   2.847                 Gaps:   3
Percent Similarity:   59.748   Percent Identity:   37.107

alignment_block:
US-09-030-606-177 x US-08-744-026-3

Align seg 1/1 to: US-08-744-026-3   from: 1   to: 258

     7  TCGCAGCCCTGGCAGGCGGCCACTGGTCATGCAGGAAAACGAATTGTTCTGCTC   56
        ||||||||||||||||||||   :::   |||::   :::   |||:
    32  SerGlnProTrpGlnAlaAlaLeuTyrHisProSerThrPheGlnLnCysGl   48

    57  GGGCGTCCTGGTCATCCGCAGTGGTGCTGGTCTGCTGTCAGCCGCACACTGTTTCC   106
        :|||::||||||||||||||||||||::||||||||||||||
    48  yGlyIleLeuValHisProGlnTrpValLeuThrAlaAlaHisCysIleG   65

   107  AGAACTCCTACACCATCGGGCTGGGCCTGCGGCCTGCACAGTCTTGAGGCCGACCAA   156
        ::::::|||     |||||:   |||||:::|||     |||::::|||
```

```
 65 lyAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAspAsp 80
    |||:|||                            |||||
157 GAGCCAGGAGACCCAGATGGTGGAGGCCAGCCTCCGTACGGCACCCAGA 206
    |||        :::|||::|||:::|||  |||   ||||||
 81 GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProCy 97

207 GTACAACAGACCCTTGCTCGCTAAC.................... 231
    :::|||   |||||  |||
 97 sPheAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAlaAspGluAspT 114

232 ......GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCT 273
    |||||||||||:::|||  :::  :::|||
114 yrSerHisAspLeuMetLeuLeuArgLeuThrGlnProAlaGluIleThr 130

274 GACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAA 323
    |||:|||:::       :::|||   |||  :::|||:::
131 AspAlaValGlnValValGluIluGluProThrGlnLeuProGluValGlySe 147

324 CTCTTGCCTCGTTTCTGGCTGGGGT.................... 348
    :::|||||||:::|||||||||
147 rThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnPheSerT 164

349 ....................CTGCTGGCGAACGATGCT 366
                       :::||| |||||
164 yrProAspLeuGlnCysValAspLeuLysIleLeuProAsnAspLys 180

367 GTGATTGCCATCCCAGACTGTG 393
    :::||||:::|||
181 CysAlaLysAlaHisThrGlnLysVal 189
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-744-026-5

```
seq_documentation_block:
; Sequence 5, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
;   APPLICANT: Bandman, Olga
;   APPLICANT: Goli, Surya K.
;   TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
;   TITLE OF INVENTION: KALLIKREIN
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;     STREET: 3174 Porter Drive
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: US
;     ZIP: 94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/744,026
;     FILING DATE: Herewith
;     CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Billings, Lucy J
;     REGISTRATION NUMBER: 36,749
;     REFERENCE/DOCKET NUMBER: PF-0154 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 415-855-0555
;     TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 261 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
```

```
alignment_scores:
    Quality:  270.00        Length:      170
    Ratio:      2.700        Gaps:          4
Percent Similarity: 58.824   Percent Identity: 34.706

alignment_block:
US-09-030-606-177 x US-08-744-026-5

Align seg 1/1   to: US-08-744-026-5   from: 1   to: 261

  4 CACTCCAGCCCTGGCAGGCGGCACTGGTCATGAGAAAAGAATTGTTCTG 53
    |||||||||||:::|||::|||  :::|||::|||:::
 34 HisSerGlnProTrpGlnValLeuValAlaSerArgGlyArgAlaValCy 50

 54 CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 103
    |:::||||||||||||||||||||||||||||:::  :::|||
 50 sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysI 67

104 TCCAGAACTCCTACCACCATCGGCGTGGGCCTGCACAGTCTTGAGGCCGAC 153
    :::|||||:::   :::|||  |||||   |||||   |||||||
 67 leArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 83

154 CAAGAGCCAGGGAGCCAGATGGTGAGGCCAGCCTCCGTACGGCACCC 203
    :::|||    |||    |||    |||  :::|||::::|||||||
 84 GluAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr 99

204 AGAGTACAACAGACCCTTGCTCGCTAAC.................... 231
    |||  |||:||     ||||||  |||
 99 oLeuTyrAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 116

232 .........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
           ||||||||||||||||||||:::||||   :::|||
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluLeu 132

271 TCTGCACCATCCGGAGCATCAGCATTGCTTCCGCAGTGCCCTACCGCGGGG 320
    :::|||:::|||   :::|||::::|||   |||||||:::   |||
133 ThrAspAlaValLysValMetAspLeuProThrGlnGluProAlaLeuGlu 149

321 GAACTCTTGCCTCGTTTCTGGCTGGGGT.................... 348
    |:::::|||   :::||||||||||||||
149 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 166

349 .........................CTGCTGGCGAACGAT 363
                            ||||||||||
166 euThrProLysLeuGlnCysValAspLeuHisValIleSerAsnAsp 182

364 GCTGTGATTGCCATCCCAGACTGTG.................... 393
    :::   ||||||   |||      |||
183 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 199

394 .GGAGGCTGG 402
     |||    |||
199 aGlyArgTrp 202
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-744-026-4

```
seq_documentation_block:
; Sequence 4, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
;   APPLICANT: Bandman, Olga
;   APPLICANT: Goli, Surya K.
;   TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
;   TITLE OF INVENTION: KALLIKREIN
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
```

```
;     STREET:   3174 Porter Drive
;     CITY:     Palo Alto
;     STATE:    CA
;     COUNTRY:  US
;     ZIP:      94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/744,026
;     FILING DATE:  Herewith
;     CLASSIFICATION:  514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Billings, Lucy J
;     REGISTRATION NUMBER:  36,749
;     REFERENCE/DOCKET NUMBER:  PF-0154 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415-855-0555
;     TELEFAX:   415-845-4166
; INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  262 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   IMMEDIATE SOURCE:
;     LIBRARY:  GenBank
;     CLONE:  186651
US-08-744-026-4

alignment_scores:
        Quality:   268.50         Length:     155
        Ratio:     2.658          Gaps:        4
Percent Similarity:  65.161   Percent Identity:  38.065

alignment_block:
US-09-030-606-177 x US-08-744-026-4

Align seg 1/1  to: US-08-744-026-4  from: 1  to: 262  .

  4 CACTCGCAGCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG  53
    ||||||||||||||||||||||||||        ::  ::  ||  ||
 34 HisSerGlnProTrpGlnAlaAlaLeuTyrHisPheSerThrPheGlnCy  50

 54 CTCGGGCGTCCTGTGCATCCGCAGTGGGTGCTGTCAGCCGCCACACTGTT 103
    I::::::|||||||  ||||||||  |||||||||:::||||||||||||
 50 sGlyGlyIleLeuValHisArgGlnTrpValLeuThrAlaAlaHisCysI  67

104 TCCAGAACTCCTACACCATCGGGCTTGGGCCTGCCACAGTCTTGAGGCCGAC 153
    :::::::::|| ||  :: ||||| ||::::||| || ||
 67 leSerAspAsnTyrGlnLeuTrpLeuGlyArgHisAsnLeu...PheAsp  82

154 CAAGAGCCAGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCCACCC 203
    :::||| ||:::||::|:::||  ||| ||||           ||||
 83 AspGluAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr  99

204 AGAGTACAACACAGACCCTTGCTCGCTAAC.................... 231
    ::|||  ||||| |||
 99 oGlyPheAsnMetSerLeuLeuGluAsnHisThrArgGlnAlaAspGluA 116

232 ..........GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
    ||||||||||||||::::::  ||| |||  :::::::
116 spTyrSerHisAspLeuMetLeuIleLysLeuThrGluProAlaAspThr 132

271 ...TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGC 317
    :::||||::::::: :: ::|||      ||| || :::
133 IleThrAspAlaValLysValValGluGluLeuProThrGlnGluProGluValVa 149
```

```
318 GGGGAACTCTTGCCTCCGTTCTGGCTGGGGTCTGCTGGCG.......... 357
    :||::|::|||||||:::||||||:||||||||
149 lGlySerThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnP 166

358 .......AACGATGCTGTGATTGCCATCCAGTCCAGACTGTGGGAGGGC 399
    |||      :::::::|||  :::  ||:  ::
166 heSerPheProAspPheGlnCysValAspLeuLysIleLeuProAsn 182

400 TGGGAGTGTGAGAAG 414
    |||||||:::|||
183 AspGluCysLysLys 187

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-8
```

```
seq_documentation_block:
; Sequence 8, Application PC/TUS9506157
; GENERAL INFORMATION:
;   APPLICANT:  Mayo Foundation for Medical Education
;   APPLICANT:  and Research
;   APPLICANT:  Hybritech Incorporated
;   APPLICANT:  Tindall, Donald J.
;   APPLICANT:  Young, Charles Y.F.
;   APPLICANT:  Saedi, Mohammed S.
;   TITLE OF INVENTION:  Recombinant HK2 Polypeptide
;   NUMBER OF SEQUENCES:  18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Schwegman, Lundberg & Woessner, P.A.
;     STREET:   3500 IDS Center
;     CITY:     Minneapolis
;     STATE:    MN
;     COUNTRY:  USA
;     ZIP:      55402
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC Compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US95/06157
;     FILING DATE:
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Raasch, Kevin W.
;     REGISTRATION NUMBER:  35,561
;     REFERENCE/DOCKET NUMBER:  150.148WO1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  612-339-0331
;     TELEFAX:   612-339-3061
; INFORMATION FOR SEQ ID NO:  8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  238 amino acids
;     TYPE:  amino acid
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
PCT-US95-06157-8

alignment_scores:
        Quality:   266.50         Length:     160
        Ratio:     2.835          Gaps:        3
Percent Similarity:  58.750   Percent Identity:  36.250

alignment_block:
US-09-030-606-177 x PCT-US95-06157-8

Align seg 1/1  to: PCT-US95-06157-8  from: 1  to: 238

  4 CACTCGCAGCCCTGGCAGGCGGCCACTGGTCATGGAAAACGAATTGTTCTG  53
    ||||||||||||||||||||:::||
 11 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy  27

 54 CTCGGGCGTCCTGTGCATCCGCAGTGGTGCTGTCAGCCGCCACACTGTT 103
```

```
                    LENGTH:  244 amino acids
                    TYPE:  amino acid
                    TOPOLOGY:  linear
                    MOLECULE TYPE:  protein
              PCT-US95-06157-10


              alignment_scores:
                    Quality:  266.50           Length:  160
                    Ratio:  2.835              Gaps:  3
              Percent Similarity:  58.750    Percent Identity:  36.250

              alignment_block:
              US-09-030-606-177 x PCT-US95-06157-10

              Align seg 1/1  to: PCT-US95-06157-10  from: 1  to: 244

                4 CACTCGCAGCCCTGGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG 53
                  ||||||||||||||||||||||||||::|||     :::::
               17 HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy 33

               54 CTCGGGCGTCCTGTGCATCCCAGTGGTGCTGTCAGCCGCACACTGTT 103
                  |:|||||||||||||||||||||||||||||:::||||||||||||
               33 sGlyGlyValLeuValHisProGlnProValLeuThrAlaAlaHisCysL 50

              104 TCCAGAACTCCTACACCATCGGGCTGCGGCCTGCCACAGTCTTGAGGCCGAC 153
                  ::|||||||||||||||||||||||||  |||::|||
               50 euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro 66

              154 CAAGAGGCCAGGGAGCCAGGATGGTGGAGGCCAGCCCTCCGTACGGCACCC 203
                  :::::||  |||||  |||  ::|||  |||    |||||
               67 GluArgVal...GlnArgValProValSerHisSerPheProHisPr 82

              204 AGAGTACAACAGACCCTTGCTC.................... 225
                  ||||||     |||||
               82 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 99

              226 .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
                       ::::||||||||||||||||||||||||   ::::|||
               99 spSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 115

              271 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCCCTACCGCGGG 320
                  ::|||::::::   :::::::::   ::|||| |||:::||
              116 ThrAspValLysValLeuGlyIleLeuProThrArgLeuProAlaLeuGl 132

              321 GAACTCTTGCCTCGTTTCTGCTGGGGGT.............. 348
                  ::|||::|||||||
              132 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 149

              349 ....................CTGCTGGCGAACGAT 363
                                      ||||||||||:::|||||
              149 euArgProArgSerLeuGlnCysValSerLeuHisLeuSerAsnAsp 165

              364 GCTGTGATTGCCATCCAGTCCCAGACTGTG 393
                  :|||:::||   |||
              166 MetCysAlaArgAlaTyrSerGluLysVal 175

              seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-6

              seq_documentation_block:
              ; Sequence 6, Application PC/TUS9506157
              ; GENERAL INFORMATION:
              ;    APPLICANT:  Mayo Foundation for Medical Education
              ;    APPLICANT:  and Research
              ;    APPLICANT:  Hybritech Incorporated
              ;    APPLICANT:  Tindall, Donald J.
              ;    APPLICANT:  Young, Charles Y.F.
              ;    APPLICANT:  Saedi, Mohammed S.
              ;    TITLE OF INVENTION:  Recombinant HK2 Polypeptide
              ;    NUMBER OF SEQUENCES:  18
              ;    CORRESPONDENCE ADDRESS:
              ;      ADDRESSEE:  Schwegman, Lundberg & Woessner, P.A.
              ;      STREET:  3500 IDS Center
```

```
                  |::::||||||||||||||||||||||||||||||||::||||||||||
               27 sGlyGlyValLeuValHisProGlnTrpProValLeuThrAlaAlaHisCysL 44

              104 TCCAGAACTCCTACACCATCGGGCTGCGGCCTGCCACAGTCTTGAGGCCGAC 153
                  ::|||||::: :: ||||||||||||| ||||:::|||
               61 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 76

              204 AGAGTACAACAGACCCTTGCTC.................... 225
                  |  |||||
               76 oLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA 93

              226 ....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG 270
                       :::::|||||||||||||||||||:::||   ::::|||
               93 spSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 109

              271 TCTGACACCATCCGGAGCATCAGCATTGCTTCGCCCTACCGCGGG 320
                  ::|||::::::   ::::::::   ::||| |||:::|
              110 ThrAspValLysValLeuGlyIleLeuProThrArgLeuProAlaLeuGl 126

              321 GAACTCTTGCCTCGTTTCTGCTGGGGGT.............. 348
                  ::||:   ::|||
              126 yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL 143

              349 ....................CTGCTGGCGAACGAT 363
                                      ||||||::::|||||
              143 euArgProArgSerLeuGlnCysValSerLeuHisLeuSerAsnAsp 159

              364 GCTGTGATTGCCATCCAGTCCCAGACTGTG 393
                  :|||::: || |||
              160 MetCysAlaArgAlaTyrSerGluLysVal 169

              seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-06157-10

              seq_documentation_block:
              ; Sequence 10, Application PC/TUS9506157
              ; GENERAL INFORMATION:
              ;    APPLICANT:  Mayo Foundation for Medical Education
              ;    APPLICANT:  and Research
              ;    APPLICANT:  Hybritech Incorporated
              ;    APPLICANT:  Tindall, Donald J.
              ;    APPLICANT:  Young, Charles Y.F.
              ;    APPLICANT:  Saedi, Mohammed S.
              ;    TITLE OF INVENTION:  Recombinant HK2 Polypeptide
              ;    NUMBER OF SEQUENCES:  18
              ;    CORRESPONDENCE ADDRESS:
              ;      ADDRESSEE:  Schwegman, Lundberg & Woessner, P.A.
              ;      STREET:  3500 IDS Center
              ;      CITY:  Minneapolis
              ;      STATE:  MN
              ;      COUNTRY:  USA
              ;      ZIP:  55402
              ; COMPUTER READABLE FORM:
              ;    MEDIUM TYPE:  Floppy disk
              ;    COMPUTER:  IBM PC compatible
              ;    OPERATING SYSTEM:  PC-DOS/MS-DOS
              ;    SOFTWARE:  PatentIn Release #1.0, Version #1.25
              ; CURRENT APPLICATION DATA:
              ;    APPLICATION NUMBER:  PCT/US95/06157
              ;    FILING DATE:
              ;    CLASSIFICATION:
              ; ATTORNEY/AGENT INFORMATION:
              ;    NAME:  Raasch, Kevin W.
              ;    REGISTRATION NUMBER:  35,561
              ;    REFERENCE/DOCKET NUMBER:  150.148WO1
              ; TELECOMMUNICATION INFORMATION:
              ;    TELEPHONE:  612-339-0331
              ;    TELEFAX:  612-339-3061
              ; INFORMATION FOR SEQ ID NO:  10:
              ;    SEQUENCE CHARACTERISTICS:
```

```
;   CITY:  Minneapolis
;   STATE:  MN
;   COUNTRY:  USA
;   ZIP:  55402
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  PCT/US95/06157
;   FILING DATE:
;   CLASSIFICATION:
;  ATTORNEY/AGENT INFORMATION:
;   NAME:  Raasch, Kevin W.
;   REGISTRATION NUMBER:  35,561
;   REFERENCE/DOCKET NUMBER:  150.148WO1
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  612-339-0331
;   TELEFAX:  612-339-3061
;  INFORMATION FOR SEQ ID NO:  6:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH:  261 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
PCT-US95-06157-6


alignment_scores:
         Quality:   266.50          Length:    160
         Ratio:     2.835           Gaps:        3
Percent Similarity:  58.750    Percent Identity:  36.250

alignment_block:
US-09-030-606-177  x  PCT-US95-06157-6

Align seg 1/1  to:  PCT-US95-06157-6  from:  1  to:  261

    4  CACTCGACAGCCCTGCAGGCGGCACTGGTCATGGAAAACGAATTGTTCTG   53
       ||||||||||||||||::||:::::      :::::
   34  HisSerGlnProTrpGlnValAlaValTyrSerHisGlyTrpAlaHisCy   50

   54  CTCGGGCGTCCTGGTGCATCCGCAGTGGGTGCTGTCAGCCGCACACTGTT  103
       |:::|||||||||||||||||||||||||::||||||||||||
   50  sGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL   67

  104  TCCAGAACTCCTACACCATCGGGCTGGGCCTGCACCAGTCTTGAGGCCGAC  153
       ::::::::::       :::||||| ||::::|||
   67  euLysLysAsnSerGlnValTrpLeuGlyArgHisAsnLeuPheGluPro   83

  154  CAAGAGCCAGGGAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCC  203
       ::::::  |||    ||| :::|| ||   ||||   |||||
   84  GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr   99

  204  AGAGTACAACAGACCCTTGCTC............................  225
       | ||||||     ||||||
   99  OLeuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGluA  116

  226  .....GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAG  270
       :::::::||||||||||::::::::|||      :::::::
  116  spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle  132

  271  TCTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGG  320
       ::||::||||| :::::::    :::||  |||:  |||::::  ||
  133  ThrAspValValLysValLeuGlyLeuProThrGlnProAlaLeuGl     149

  321  GAACTCTTGCCTCGTTCTGGCTGGGT........................  348
       |::::||| ::||||||||||||
  149  yThrThrCysTyrAlaSerGlyTrpGlySerIleGluProGluGluPheL  166

  349  ...............................CTGCTGGCGAACGAT      363
                                      |||||::::|||||
```

```
  166  euArgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerAsnAsp  182

  364  GCTGTGGATTGCCATCCAGTCCCAGACTGTG  393
                    ||:::::       ||
  183  MetCysAlaArgAlaTyrSerGluLysVal  192
```

Wed Sep 29 14:27:33 1999

OM of: US-09-030-606-174 to: A_Geneseq_36:*   out_format : pfs
Date: Sep 25, 1999 11:35 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL-frame+ n2p.model -DEV-xlp
 -Q-/cgn2_1/USPTO_spool/US09030606/runat_24091999_171616_29804/app_query.fasta.1
 -DB-A_Geneseq_36 -QFMT-fastan -SUFFIX-rag -GAPOP-12.000
 -GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
 -QGAPOP-4.500 -QGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.000
 -FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
 -DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-blosum62
 -TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR_SCORE-escore
 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-stat -USER-US09030606
 -NCPU-6 -ICPU-3 -WAIT -THREADS-1

Search information block:
Query: US-09-030-606-174
Query length: 1459
Database: A_Geneseq_36:*
Database sequences: 188963
Database length: 2368106
Search time (sec): 185.540000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:W59129	+	672.00	1103.83	2.3e-54	232	Homo sapiens Tub Interactor (hT
A_Geneseq_36:W69388	+	626.00	1028.54	4.0e-50	205	Prostate tumour specific gene c
A_Geneseq_36:W71872	+	626.00	1028.54	4.0e-50	205	Protein encoded by prostate tum
A_Geneseq_36:W60592	+	623.00	1021.62	8.0e-50	248	Human prostate-specific kallikr
A_Geneseq_36:W69387	+	502.00	824.77	1.1e-38	159	Prostate tumour specific gene c
A_Geneseq_36:W71871	+	502.00	824.77	1.1e-38	159	Protein encoded by prostate tum
A_Geneseq_36:W69389	+	462.00	757.89	5.9e-35	164	Prostate tumour specific gene c
A_Geneseq_36:W71873	+	462.00	757.89	5.9e-35	164	Protein encoded by prostate tum
A_Geneseq_36:Y12281	+	326.00	540.80	1.8e-22	66	Human 5' EST secreted protein SE
A_Geneseq_36:R95913	-	292.50	466.88	4.0e-19	397	Neural thread protein. Detectio
A_Geneseq_36:W81504	+	280.00	435.95	7.7e-18	1079	Short form of TPR motif Y (TPR
A_Geneseq_36:W88627	-	263.50	434.00	1.2e-16	87	Secreted protein encoded by gene
A_Geneseq_36:W94493	+	256.00	410.12	8.5e-16	268	Human kallikrein. Human kallikr
A_Geneseq_36:Y02693	-	249.00	407.49	2.9e-15	110	Human secreted protein encoded
A_Geneseq_36:W48351	+	249.00	406.36	3.0e-15	123	Human breast cancer related pro
A_Geneseq_36:R13556	+	247.00	398.31	5.3e-15	196	Protein encoded downstream of H
A_Geneseq_36:Y02785	+	245.50	400.62	6.3e-15	122	Human secreted protein encoded
A_Geneseq_36:W40353	+	242.00	389.99	1.5e-14	196	Human unspecified protein from
A_Geneseq_36:R07057	+	242.00	389.20	1.6e-14	212	Smaller hepatocellular oncoprot
A_Geneseq_36:R53638	+	227.50	364.24	3.6e-13	230	Bovine trypsinogen. Expression
A_Geneseq_36:W81767	+	226.50	362.89	4.4e-13	223	Bovine TRYP peptide fragment. D
A_Geneseq_36:R53637	+	226.50	362.85	4.4e-13	224	Bovine trypsin. Expression vect
A_Geneseq_36:R67888	+	222.00	354.12	1.2e-12	253	Human stratum corneum chymotryp
A_Geneseq_36:W05383	+	222.00	354.12	1.2e-12	253	Human amyloid precursor protein
A_Geneseq_36:R82703	+	221.50	354.26	1.3e-12	230	Human TRYI trypsinogen variant
A_Geneseq_36:W57740	+	211.50	337.19	1.1e-11	240	Trypsinogen-like protein. DNA c
A_Geneseq_36:W64260	+	208.50	331.94	2.1e-11	246	Human amyloid beta-protein prec
A_Geneseq_36:W08475	+	207.50	330.24	2.6e-11	247	Porcine trypsinogen. Prodn. of
A_Geneseq_36:W10694	+	205.50	326.39	4.0e-11	260	Human recombinant neuropsin, us
A_Geneseq_36:W12393	+	205.50	326.39	4.0e-11	260	Mouse neuropsin protein. Nuclei
A_Geneseq_36:W00838	-	200.00	314.70	1.4e-10	334	Tumour necrosis factor-related
A_Geneseq_36:Y02925	-	194.00	322.81	3.0e-10	56	Fragment of human secreted prote
A_Geneseq_36:P81243	+	194.50	308.61	4.2e-10	247	Human spleen trypsin III (trysi
A_Geneseq_36:R82703	+	194.50	308.61	4.2e-10	247	Human pancreatic trypsin III. N
A_Geneseq_36:R84667	+	191.50	304.03	7.8e-10	237	Mature kallikrein hK2. New isol
A_Geneseq_36:W45395	+	191.50	304.03	7.8e-10	237	Mature prostate-specific glandu
A_Geneseq_36:W49087	+	191.50	304.03	7.8e-10	237	Mutant human Kallikrein 2 (hK2
A_Geneseq_36:W83212	+	191.50	304.03	7.8e-10	237	hK2 variant A217V. Detection of
A_Geneseq_36:W83202	+	191.50	304.03	7.8e-10	237	Prostate-specific glandular kal
A_Geneseq_36:W96186	+	191.50	304.03	7.8e-10	237	Mature human Kallikrein 2 (hK2
A_Geneseq_36:W96190	+	191.50	304.03	7.8e-10	237	Variant human Kallikrein 2 (hK2
A_Geneseq_36:R84670	+	191.50	303.99	7.8e-10	238	Mature kallikrein hK2. New isol

A_Geneseq_36:R84669	+	191.50	303.74	7.9e-10	244	Pro-hK2 kallikrein. New isol
A_Geneseq_36:W45396	+	191.50	303.74	7.9e-10	244	Prostate-specific glandular
A_Geneseq_36:W83204	+	191.50	303.74	7.9e-10	244	Prostate-specific glandular

seq_name: A_Geneseq_36:W59129

seq_documentation_block:
ID   W59129 standard; Protein; 232 AA.
AC   W59129;
DT   11-SEP-1998 (first entry)
DE   Homo sapiens Tub Interactor (hTI-1) protein.
KW   serine protease; tub interactor; treatment; obesity; cachexia;
KW   anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW   neurodegenerative disease; Alzheimer's disease; drug screening;
KW   Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW   amyotrophic lateral sclerosis; spinocerebellar degeneration.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   Region   42
FT   /note= "undefined amino acid"
PN   WO9812302-A1.
PD   26-MAR-1998.
PF   05-SEP-1997; U15627.
PR   21-JUL-1997; US-897340.
PR   17-SEP-1996; US-715032.
PA   (MILL-) MILLENNIUM PHARM INC.
PI   Errada PR, Gimeno CJ;
DR   WPI; 98-217246/19.
DR   N-PSDB; V11855.
PT   Tub interactor genes   - used to develop products for the treatment
PT   of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT   diabetes
PS   Claim 28; Fig 1; 120pp; English.
CC   The sequence is that encoding the Tub interactor protein (hTI-1)
CC   which is a putative serine protease. TI genes function
CC   in biochemical pathways involved in weight control and
CC   related disorders. The products can be used for treating
CC   weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC   or a related disorder such as diabetes. The products can
CC   also be used to modulate cell cycle progression and apoptosis.
CC   They can be used for treating neurodegenerative diseases
CC   which are characterised by apoptosis, including Alzheimer's
CC   disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC   lateral sclerosis or spinocerebellar degenerations. The
CC   products can also be used for detection, diagnosis and
CC   drug screening.
SQ   Sequence 232 AA;

alignment_scores:
     Quality: 672.00   Length: 187
     Ratio: 4.541   Gaps: 3
Percent Similarity: 79.144   Percent Identity: 74.866

alignment_block:
US-09-030-606-174 x W59129

Align seg 1/1 to: W59129 from: 1 to: 232

    3 TCAGCCCACACTGTTTCCAGAAGTGAGTGCAGAGCTCCACACCATCGG   52
      ||||||||||||||||||||||||||||||||||||||||||||||||
   34 SerAlaAlaHisCysPheGlnLys***ValGlnSerSerTyrThrIleGl   50

   53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGG   102
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   50 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV   67

  103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC   152
      |||||||||||||||||||||||||||||||||||||||||||||||||||
   67 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu   83

  153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA   202
      ||||||||||||||||||||||||||||||||||||||||||||||||||

```
alignment_scores:
             Quality:  626.00         Length:      187
               Ratio:    4.378          Gaps:         4
Percent Similarity:  76.471  Percent Identity: 71.123

alignment_block:
US-09-030-606-174 x W69388

Align seg 1/1  to: W69388  from: 1  to: 205

    3 TCAGCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCCG  52
      ||||||||||||||||||      ||::||||||||||||||
   19 SerAlaAlaHisCysPhe.........GlnAsnSerTyrThrIleGl  31

   53 GCTGGGCTGCACAGTCTTGAGCGCCACCAAGAGCCAGGAGCCAGATGG 102
      ||||||||||||||||||||||||||||||:::||||||||||||||
   31 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV  48

  103 TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
      ||||||||||||||||||||:::||||||||||||||||||||||
   48 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgLeuLeuLeu  64

  153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCTCCGAGTCTGA 202
      |||||||||:::||||||||||||||||||||||||||||||||||||||
   65 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerArg  81

  203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGGGACT 252
      ||||||||||||||||||||||||||||||||||||||||||||||||
   81 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsn  98

  253 CTTGCCTCGTTCTGGCTGGGGTCTGCTGGCGAACGGTGAGCTCACGGGT 302
      ||||||||||||||||||||||||||||||||||||||
   98 erCysLeuValSerGlyTrpGlyLeuLeu.................... 107

  303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGAC 352
      ...........................:::::::::::::::
  107 ................................................. 107

  353 CCAGAGCTCTGCGTCCCAGGCAGAAATGCCTACCGTGCTGCAGTGCGTGAA 402
      ||||::  ||||||||||||||||||||||:::||||||||
  107 ........AlaAsn..GlyArgMetProThrValLeuHisCysValAs 120

  403 CGTGTCGGTGGTGTCTGANGAGGTCGCANTAAGCTCATGACCCGCTGT 452
      ||||||||||||||::::|||||:::|||||||||||
  120 nValSerValValSerGlu***ValCysSerLysLeuTyrAspProLeuT 137

  453 ACCACCCCANCATGTTCTCGCGCCGGGAGGGGAGGCAAGACCAGAAGGACTCC 502
      |||||||:::|||||||||||||||||||||||||||||||||||||
  137 yrHisProSerMetPheCysAlaGlyGlyGlyProLeuIleCysAsnGlyTy 153

  503 TGCAACGTGAGAGAGGGGAAAGGGGAGGGGAGGGGCGGACTCAGGGAAGGGTG 552
      ||||||      |||  ||:  ::::|||||||||||
  154 CysAsn.......GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 167

  553 GAGAAGGGGG 562
      : ::||||
  167 rLeuGlnGly 170
```

seq_name: A_Geneseq_36:W71872

```
seq_documentation_block:
ID   W71872 standard; Protein; 205 AA.
AC   W71872;
DT   06-JAN-1999  (first entry)
DE   Protein encoded by prostate tumour clone P703 splice variant DE13.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 127
FT                   /note= "undefined residue"
FT   Misc_difference 204
FT                   /note= "undefined residue"
```

[The following is the left column, earlier alignment against W69388:]

```
   84 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs 100

  203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGGGACT 252
      ||||||||||||||||||||||||||||||||||||||||||||||||
  100 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsn 117

  253 CTTGCCTCGTTCTGGCTGGGGTCTGCTGGCGAACGGTGAGCTCACGGGT 302
      ||||||||||||||||||||||||||||||||||||||
  117 erCysLeuValSerGlyTrpGlyLeuLeu.................... 126

  303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGAC 352
      ...................................................
  126 ................................................. 126

  353 CCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
      ||||::  ||||||||||||||||||||||:::||||||||
  127 ........AlaAsn..GlyArgMetProThrValLeuGlnCysValAs 139

  403 CGTGTCGGTGGTGTCTGANGAGGTCGCANTAAGCTCATGACCCGCTGT 452
      ||||||||||||||::::|||||:::|||||||||||
  139 nValSerValValSerGluGluValCysSerLysLeuTyrAspProLeuT 156

  453 ACCACCCCANCATGTTCTCGCGCCGGGAGGGGAGGCAAGACCAGAAGGACTCC 502
      |||||||:::|||||||||||||||||||||||||||||||||||
  156 yrHisProSerMetPheCysAlaGlyGlyGlyProLeuIleCysAsnGlyTy 172

  503 TGCAACGTGAGAGAGGGGAAAGGGGAGGGGAGGGGCGGACTCAGGGAAGGGTG 552
      ||||||      |||  ||:  ::::|||||||||||
  173 CysAsn.......GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 186

  553 GAGAAGGGGG 562
      : ::||||
  186 rLeuGlnGly 189
```

seq_name: A_Geneseq_36:W69388

```
seq_documentation_block:
ID   W69388 standard; Protein; 205 AA.
AC   W69388;
DT   08-DEC-1998  (first entry)
DE   Prostate tumour specific gene clone DE13 protein.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 127
FT                   /note= "unspecified amino acid"
FT   Misc_difference 204
FT                   /note= "unspecified amino acid"
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   N-PSDB; V58647.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Example 1; Page 115-116; 141pp; English.
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence   205 AA;
```

```
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03492.
PR   09-FEB-1998;  US-020956.
PR   25-FEB-1997;  US-806099.
PR   01-AUG-1997;  US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609986/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 3; Page 107-108; 130pp; English.
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence    205 AA;

alignment_scores:
         Quality:   626.00                 Length:      187
           Ratio:    4.378                   Gaps:        4
Percent Similarity:   76.471    Percent Identity: 71.123

alignment_block:
US-09-030-606-174 x W71872

Align seq 1/1  to: W71872  from: 1  to: 205

  3 TCAGCCGCACACTGTTTCCAGAGAAGTGAGTGCAGAGCTCCTACACCATCGG   52
    ||||||||||||||||||        ||::::||||||||||||||||||||
 19 SerAlaAlaHisCysPhe..........GlnAsnSerTyrThrIleGl   31

 53 GCTGGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGGAGCCAGATGG  102
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 31 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV   48

103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC  152
    |||||||||||||||||||||||||||||||||||||  |||||||||||
 48 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgLeuLeuLeu   64

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA  202
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 65 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs   81

203 CACCATCCGGAGCATCAGCATTGCTTCGCCAGTGCCCTACCGCGGGGAACT  252
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 81 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnS   98

253 CTTGCCTCCGTTCGTCGGGGTCTGCTGGGCCGAACGGTGAGCGTCACGGGT  302
    ||||||||||||:::||||||||||||::::||||||||||||||||
 98 erCysLeuValSerGlyTrpGlyLeuLeu...................  107

303 GTGTGTCTGCCCCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGAC   352
 107 ..................................................  107

353 CCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCCAGTGCGTGAA  402
    |||:::  |||||||||||||||||||||::::||||||||
108 ........AlaAsn..GlyArgMetProThrValLeuHisCysValAs  120

403 CGTGTCGGTGGTGTCTGANGAGGTCTCANTAAGCTCTATGACCCGCTGT  452
    ||||||||||||||||    ||||:::||||||||||||||||||||
120 nValSerValValSerGlu***ValCysSerLysLeuTyrAspProLeuT  137

453 ACACCCCANCANCATGTTCTGCGCCCGGCCGAGGGCAAGACCAGAAGGACTCC  502
    ||||:::||||||||||||||||||||||||||||||||||||||||||
137 yrHisProSerMetPheCysAlaGlyGlyGlyGlnLysAspAspSer  153

503 TGCAACGTGAGAGGGGGAAAGGGGAGGGCAGGGCCACTCCAGGGAAGGGTG  552
    ||||||||        |||        :::  ::||||:::
```

seq_name: A_Geneseq_36:W60592

```
154 CysAsn........GlyAspSerGlyGlyProLeuIleCysAsnGlyTy  167

553 GAGAAGGGGG  562
    :  ::::||
167 rLeuGlnGly  170
```

seq_documentation_block:
```
ID   W60592 standard; Protein; 248 AA.
AC   W60592;
DT   07-SEP-1998  (first entry)
DE   Human prostate-specific kallikrein (HPSK) protein.
KW   Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW   benign prostate hyperplasia; diagnosis; drug screening; PSK.
OS   Homo sapiens.
FH   Key           Location/Qualifiers
FT   Misc_difference 113
FT                  /label= unknown
FT                  /note= "encoded by NTC"
FT   Misc_difference 128
FT                  /label= unknown
FT                  /note= "encoded by AGN"
FT   Misc_difference 132
FT                  /label= unknown
FT                  /note= "encoded by GNT"
PN   WO9820117-A1.
PD   14-MAY-1998.
PF   31-OCT-1997;  U20051.
PR   05-NOV-1996;  US-744026.
PA   (INCY-) INCYTE PHARM INC.
PI   Bandman O, Goli SK;
DR   WPI; 98-286933/25.
DR   N-PSDB; V37495.
PT   New isolated prostate-specific kallikrein - used to develop products
PT   for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT   hyperplasia
PS   Claim 1; Fig 1A-C; 68pp; English.
CC   This represents a human prostate-specific kallikrein (HPSK). A host cell
CC   containing an expression vector comprising the HPSK nucleic acid sequence
CC   can be used to produce the protein recombinantly. The HPSK products can
CC   be used for the diagnosis of conditions or diseases associated with
CC   expression of HPSK such as prostate carcinoma and benign prostate
CC   hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC   modulate its activity can be used for the preparation of treatment of
CC   such conditions or diseases. The products can also be used for detection
CC   and drug screening, especially for the detection of prostate-specific
CC   kallikrein (PSK).
SQ   Sequence    248 AA;

alignment_scores:
         Quality:   623.00                 Length:      187
           Ratio:    4.357                   Gaps:        4
Percent Similarity:   76.471    Percent Identity: 70.053

alignment_block:
US-09-030-606-174 x W60592

Align seq 1/1  to: W60592  from: 1  to: 248

  3 TCAGCCGCACACTGTTCCAGAAGTGAGTGCAGAGCTCCTACACCATCGG   52
    |||||||||||||||        |||:::||||||||||||||||||
 62 SerAlaAlaHisCysPhe...........GlnAsnSerTyrThrIleGl   74

 53 GCTGGGGCCTGCACAGTCTTGAGGCCCGACCAAGAGCCAGGGAGCCAGATGG  102
    |||||||||||||||||||||||||||||||||||||||||||||||||
 74 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV   91

103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC  152
    |||||||||||||||||||||||||||||||||||||||||||||
 91 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu  107
```

```
153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCTCCGAGTCTGA 202
    |||||||||||||| ||||||||||||||||||||||||||||||||||
108 AlaAsnAspLeuMet***IleLysLeuAspGluSerValSerGluSerAs 124

203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGAACT 252
    |::||||||||||::||||||||||||||||||||||||||||||||||
124 pAsnIleArg***IleSerIle***SerGlnCysProThrAlaGlyAsnP 141

253 CTTGCCTCGTTTCTGCGGTCTGGGTCTGCTGCGAACGGTGAGCTCACGGGT 302
    |||||||||||||||||||||||
141 heCysLeuValSerGlyTrpGlyLeuLeu.............. 150

303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGAC 352
150 ................................................ 150

353 CCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
    |||::: ||||||||||||||||||||||||||||
151 .........AlaAsn..GlyArgMetProThrValLeuGlnCysValAs 163

403 CGTGTCGGTGTGTCTGANGAGGTCTGCANTAAGCCTCTATGACCCCGCTGT 452
    ||||||||||||||::||||||||||| :||||||||||||
163 nValSerValValSerGluGluValCysSerLysLeuTyrAspProLeuT 180

453 ACCACCCCANCATGTTCTGCGCCGGCGAGGGCAAGACCAGAAGGACTCC 502
    |||||||||||||||||||||||||||||||||||||||||
180 yrHisProSerMetPheCysAlaGlyGlyGlyLysGlnLysAspSer 196

503 TGCAACGTGAGAGAGGGGAAAGGGAGGCAGGCGACTCAGGGAAGGGTG 552
    |||| ||| ::: :::||::
197 CysAsn........GlyAspSerGlyGlyProLeuIleCysAsnGlyTy 210

553 GAGAAGGGGG 562
    ::::|||
210 rLeuGlnGly 213
```

seq_name: A_Geneseq_36:W69387

seq_documentation_block:
```
ID   W69387 standard; Protein; 159 AA.
AC   W69387;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone DE1 protein.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   Misc_difference  103
FT                    /note= "unspecified amino acid"
FT   Misc_difference  105
FT                    /note= "unspecified amino acid"
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI') CORIXA CORP.
PI   Dillon DC; Xu J;
DR   WPI; 98-480805/41.
DR   N-PSDB; V58644.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Example 1; Page 112-113; 14lpp; English.
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancer.
```

SQ   Sequence   159 AA;

alignment_scores:
```
              Quality: 502.00        Length:   155
              Ratio:     4.365       Gaps:       3
Percent Similarity: 74.194    Percent Identity: 69.032
```

alignment_block:
US-09-030-606-174 x W69387

Align seg 1/1 to: W69387 from: 1 to: 159

```
 99 ATGGTGGAGGCCAGCCTCCGTACGGCACCCAGAGTACAACAGACCCTT 148
    ||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLe 17

149 GCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCAGT 198
    |||||||||||||||||||||||||||||||||||||||||||||||||
 17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGlus 34

199 CTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGGCGGGG 248
    ||||||||||||||||||||||||||||||||||||||||||||||||
 34 erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50

249 AACTCTTGCCTCGTTTCTGCGGTGGGGTCTGCTGCGGAACGGTGAGCTCAC 298
    |||||||||||||||||||||||||||||
 51 AsnSerCysLeuValSerGlyTrpGlyLeuLeu.............. 61

299 GGGTGTGTGTCTGCCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGC 348
 61 ................................................ 61

349 TGACCCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCG 398
    |||::: ||||||||||||||||||||||||
 62 ..........AlaAsn..GlyArgMetProThrValLeuGlnCysV 73

399 TGAACGTGTCGGTGTGTCTGANGAGGTCTGCANTAAGCCTCTATGACCCCG 448
    ||||||||||||||||::|||||||||| :::||||||||||||
 73 alAsnValSerValValSerGluGluValCysSerLysLeuTyrAspPro 89

449 CTGTACCACCCCANCATGTTCTGCGCCGGCGAGGGCAAGACCAGAAGGA 498
    ||||||||||||::|||||||||||||||||||| ||| ||
 90 LeuTyrHisProSerMetPheCysAlaGlyGlyGlyLysGln***As 106

499 CTCCTGCAACGTGAGAGAGGGGAAAGGGAGGCAGGCGACTCAGGGAAG 548
    ||||||||||         ||| ::: ||      ::::||
106 pSerCysAsn........GlyAspSerGlyGlyProLeuIleCysAsnG 120

549 GGTGGAGAAGGGGG 562
    ||:::  ::|||
120 lyTyrLeuGlnGly 124
```

seq_name: A_Geneseq_36:W71871

seq_documentation_block:
```
ID   W71871 standard; Protein; 159 AA.
AC   W71871;
DT   06-JAN-1999 (first entry)
DE   Protein encoded by prostate tumour clone P703 splice variant DE1.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   Misc_difference  103
FT                    /note= "undefined residue"
FT   Misc_difference  105
FT                    /note= "undefined residue"
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
```

```
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 3; Page 105; 130pp; English.
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence    159 AA;

alignment_scores:
        Quality:    502.00              Length:     155
        Ratio:      4.365               Gaps:       3
Percent Similarity:  74.194     Percent Identity:  69.032

alignment_block:
US-09-030-606-174 x W71871        .

Align seg 1/1   to: W71871   from: 1   to: 159

 99  ATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTT 148
     |||||||||||||||||||||||||||||||||||||||||||||||||
  1  MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLe  17

149  GCTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGT 198
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 17  uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS  34

199  CTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGG 248
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 34  erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly  50

249  AACTCTTGCCTGCTTCTCGGCTGGGGTCTCTGGCGAACGCTGAGCTCAC  298
     |||||||||||||||||||||||||||||||||||||||
 51  AsnSerCysLeuLeuLeuGlyTrpGlyLeuLeu.............      61

299  GGGTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGC  348
     ||||||||||||||||||||||||||||||    |||||||||||||||||
 61  ..................................................  61

349  TGACCCAGAGCTCTGCGTCCCAGGCAGAATGCCTACCGTGCTGCAGTGCG 398
     ||:::  ||||||||||||||||| |||||||||||||
 62  .........AlaAsn..GlyArgMetProThrValLeuGlnCysV  73

399  TGAACGTGTCGGTGTGTCTGANGAGGTCTGCANTAAGCTCTATGACCCG 448
     |||||||||||||||   |||||||||   |:|||||||||||||
 73  alAsnValSerValCysSerGluValCysSerLysLeuTyrAspPro   89

449  CTGTACCCACCCCANCATGTTCTGCGCCGGCGAGGGCAAGACCAGAAGGA 498
     ||||||||||  :::||||||||||||||||| ||| ||
 90  LeuTyrHisProSerMetPheCysAlaGlyGlyGlyGln***Gln***As 106

499  CTCCTGCAACGTGAGAGAGGGGAAAGGGGGACTCAGGGAAG 548
     |||||||||      |||  ::: |||  :::|
106  pSerCysAsn........GlyAspSerGlyGlyProLeuIleCysAsnG  120

549  GGTGAGAGGGGG 562
     ||:::  ::::||
120  lyTyrLeuGlnGly 124

seq_name: A_Geneseq_36:W69389

seq_documentation_block:
ID   W69389 standard; Protein: 164 AA.
AC   W69389;
DT   08-DEC-1998  (first entry)
DE   Prostate tumour specific gene clone DE14 protein.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
```

```
KW   therapy.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   Misc_difference 118
FT                   /note= "unspecified amino acid"
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   N-PSDB; V58648.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Example 1; Page 117-118; 141pp; English.
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence    164 AA;

alignment_scores:
        Quality:    462.00              Length:     151
        Ratio:      4.053               Gaps:       4
Percent Similarity:  75.497     Percent Identity:  67.550

alignment_block:
US-09-030-606-174 x W69389        ..

Align seg 1/1   to: W69389   from: 1   to: 164

  3  TCAGCCGCCACACTGTTCCAGAAGTGAGTGCAGAGTCCTCACACCATCGG  52
     ||||||||||||||||             |||:::||||||||||||||
 19  SerAlaAlaHisCysPhe.............GlnAsnSerTyrThrIleGl  31

 53  GCTGGGCCTGCACAGTCTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGG 102
     |||||||||||||||||||||||||||||||||||||||||||||||
 31  yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV  48

103  TGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 48  alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu  64

153  GCTAACGACCTCATCGTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
     |||||||||||||||||||||||||||||||||||||||||||||||||
 65  AlaAsnAspLeuMetLeuIleLeuLysLeuAspGluSerValSerSerAs  81

203  CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGGAACT 252
     |||||||||||||||||||||||||||||||||||||||||||||||||
 81  pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnS  98

253  CTTGCCTGCTTCTCGGCTGGGGTCTCTGGCGAACGGTGAGCTCACGGGT 302
     ||||||||||||||||||||||||||||||||||||||||     :::
 98  erCysLeuValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAla 114

303  GTG..........................TGTCTGCCCCTTCAAGGAG 325
     :::                                      |||
115  IleGlnSer***ThrValGlyGlyTrpGluCys............GluLy 127

326  GTCCTCTGCCCAGTCGCGGGGGCTGACCCAGAGCTCTGCGTCCCAGGCAG 375
     :  |||   ::::||| |||  |||:::||::::|||:
127  sLeuSerGlnProTrpGlnGlyCysThrIleSerAlaThrSerAlaAlaA 144

376  AATGCCTACCGTGCTGCAGTGCGTGAACGTGCGGTGGTGTCTGANGAGG 425
```

```
              ::           ||||||                             |||:::
144 rg...ThrSerCysCysIleLeuThrGlyCysSerLeuLeuThrAla 159

426 TCT 428
    |||
160 Ser 160

seq_name: A_Geneseq_36:W71873

seq_documentation_block:
ID   W71873 standard; Protein; 164 AA.
AC   W71873;
DT   06-JAN-1999 (first entry)
DE   Protein encoded by prostate tumour clone P703 splice variant DE14.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
FH   Key           Location/Qualifiers
FT   Misc_difference 118
FT                  /note= "undefined residue"
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1998; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 3; Page 109; 130pp; English.
CC   The Present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence  164 AA;

alignment_scores:
     Quality:    462.00           Length:    151
     Ratio:      4.053            Gaps:      4
Percent Similarity:  75.497    Percent Identity:  67.550

alignment_block:
US-09-030-606-174 x W71873       ..

Align seg 1/1  to: W71873  from: 1  to: 164

  3 TCAGCCGCACACTGTTTCCAGAAGTGAGTGAGAGAGCTCCTACACCATCGG 52
    ||||||||||||||          ||:::::||||:||||||||||||||||
 19 SerAlaAlaHisCysPhe..........GlnAsnSerTyrThrIleGl 31

 53 GCTGGGGCCTGCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
    |||||||||||||||||||||||||||||||||||||||||||||||||
 31 yLeuGlyLeuHisSerLeuGluAlaAspGlnGluProGlySerGlnMetV 48

103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCCTTGCTC 152
    |||||||||||||||||||||||||||||||||||||||||||||||||
 48 alGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeuLeu 64

153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 65 AlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluSerAs 81

203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCCGCGGGAACT 252
    |||||||||||||||||||||||||||||||||||||||||||||||||
 81 pThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnS 98

253 CTTGCCTCGTTTCTGGCTGGGGTCTGCTGGCGAACGGTGAGCTCACGGGT 302
    ||||||||                                    ||::::
 98 erCysSerLeuValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAla 114
```

```
303 GTG.................................TGTCTGCCCTCTTCAAGGAG 325
    |||                                     :::::
115 IleGlnSer***ThrValGlyGlyTrpGluCys..............GluLy 127

326 GTCCTCTGCCCAGTCGCGGGGGCTGACCCAGAGCTCTGCGTCCAGGCAG 375
    : |||     ||     ||||||||||||:||||:||||:||||:::||:
127 sLeuSerGlnProTrpGlnGlyCysThrIleSerAlaThrSerSerAlaA 144

376 AATGCTACCGTGCTGCAGTGCGTGAACGTGTCGGTGGTCTGANGAGG 425
    ::  |||||                 |||     |||:||::
144 rg...ThrSerCysCysIleLeuThrGlyCysSerLeuLeuThrArg 159

426 TCT 428
    |||
160 Ser 160

seq_name: A_Geneseq_36:Y12281

seq_documentation_block:
ID   Y12281 standard; Protein; 66 AA.
AC   Y12281;
DT   17-JUN-1999 (first entry)
DE   Human 5' EST secreted protein SEQ ID NO:312.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
OS   Homo sapiens.
PN   WO9905548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   N-PSDB; X41114.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 27; Page 661; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence  66 AA;

alignment_scores:
     Quality:    326.00           Length:    65
     Ratio:      5.015            Gaps:      0
Percent Similarity:  100.000    Percent Identity:  98.462

alignment_block:
US-09-030-606-174 x Y12281       ..

Align seg 1/1  to: Y12281  from: 1  to: 66
```

```
99  ATGGTGGGAGGCCAGCCTCCCGTACGGCACCCAGAGTACAACAGACCCTT  148
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1   MetValGluAlaSerLeuuSerValArgHisProGluTyrAsnArgProLe  17

149 GCTCCCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGT  198
    ||||||||||||||||||||||||||||||||||||||||||||||||
17  uLeuAlaAsnAspLeuMetLeuIleLysGluSerValSerGlus  34

199 CTGACACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCCTACCGCGGGG  248
    ||||||||||||||||||||||||||||||||||||||||||||||||
34  erAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly  50

249 AACTCTTGCCTCGTTCTGGCTGGGTGGTCTGCTGGCGAACGGTGAG  293
    ||||||||||||||||||||||||||||||||||||||||||||||
51  AsnSerCysLeuValSerGlyTrpGlyLeuLeuAlaAsnGlyGln  65
```

seq_name: A_Geneseq_36:R95913

seq_documentation_block:
```
ID  R95913 standard; Protein; 397 AA.
AC  R95913;
DT  13-NOV-1996 (first entry)
DE  Neural thread protein.
KW  Neural thread protein; NTP; diagnosis; detection;
KW  Alzheimer's disease; neuroectodermal tumour; malignant astrocytoma;
KW  monoclonal antibody; binding fragment.
OS  Homo sapiens.
PN  WO961522-A1.
PD  23-MAY-1996.
PF  14-NOV-1995; U17111.
PR  14-NOV-1994; US-340426.
PA  (GEHO ) GEN HOSPITAL CORP.
PI  De LA MONTE S, Wands JR;
DR  WPI; 96-259865/26.
DR  N-PSDB; T27738.
PT  Detection of neural thread protein in diagnosis of Alzheimer's
PT  disease - also NTP DNA and protein sequences used in gene and
PT  anti-sense therapy
PS  Claim 22; Page 171-172; 238pp; English.
CC  A method for detecting the presence of neural thread protein (NTP)
CC  having a molecular weight of 8, 14, 17, 21, 26 or 42 kD in a human
CC  subject comprises (a) contacting a sample from a human subject that
CC  is suspected of containing the NTP with at least one molecule
CC  capable of binding to the protein; and (b) detecting any of the
CC  molecule bound to the protein. The binding molecule is selected
CC  from an antibody free of natural impurities, a monoclonal antibody
CC  or a binding fragment of either of these. The method may be used for
CC  diagnosing the presence of Alzheimer's disease, neuroectodermal
CC  tumours and a malignant astrocytoma in a human.
SQ  Sequence 397 AA;
```

alignment_scores:
```
        Quality:   292.50              Length:    100
        Ratio:       3.611             Gaps:        2
Percent Similarity: 81.000      Percent Identity: 66.000
```

alignment_block:
US-09-030-606-174/rev-x R95913

Align seg 1/1 to: R95913 from: 1 to: 397

```
1459 TTTTTTTTTTTTTTTTGAGACAGAGTCTTACTCGTTGCCCAGGCTGG  1410
     |||  |||  :::||||||    |||||||:::||||:::|||||||
298  PheAsnPheCysLeuPheGluMetGluSerHisSerValThrGlnAlaGl  314

1409 AGTATAGTGGTGTGATCTCAACTCACTCAACCTCTGCCTCCCATATTCA  1360
     ||||  |||  :::||||||||||||:||  :::||||||:::||||||
314  yValGlnTrpProAsnLeuGlySerLeuGlnProLeuProProGlyLeuL  331

1359 AGCAATTCCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCTG  1310
     ||  :::||||||||||||||||||||||||||||||:::||||||  |||
331  ysArgPheSerCysLeuSerLeuProSerSerTrpAspTyr.GlyHisLe  347
```

```
1309 CCACCATATCCAGCTAACTTTGTATTTTAGTACAGACAGGATTCACC  1260
     ||||| || |||||| ||||  :::::::||||||||||||||||||
347  uHisHisThrProLeuIlePheValPheSerLeuGluAlaGlyPheHisH  364

1259 ATTTTGGCCAGGCTGCTGTCTTGAACTCCTTACCTCAAGTGATCTGCCTGCC  1210
     ||:::   ||||||||:||||||||||||||||||||||:::|||||||
364  isIleCysGlnAlaGlyLeuLysLeuThrSerGlyAspProProAla  380

1209 TCGCCTCC.CAAAGTGCTGGGATTACAGGCATGAGC...CACCCTGCC  1166
     |||  ||||||||||||||||||||:::::::  ||||||
381  SerAlaPheGlnSerAlaGlyLyIleThrGlyValThrProHisProAla  396
```

seq_name: A_Geneseq_36:W81504

seq_documentation_block:
```
ID  W81504 standard; Protein; 1079 AA.
AC  W81504;
DT  02-FEB-1999 (first entry)
DE  Short form of TPR motif Y (TPRY) gene product.
KW  Non-recombining region; human; Y chromosome; X homologue; testis; TPRY;
KW  infertility; sperm; gene alteration; inhibitor.
OS  Homo sapiens.
PN  WO984674-A2.
PD  22-OCT-1998.
PF  10-APR-1998; U07115.
PR  11-APR-1997; US-041877.
PA  (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PI  Lahn BT, Page DC;
DR  WPI; 98-558729/48.
DR  N-PSDB; V69634.
PT  Novel genes in the non-combining region of Y chromosome - useful to
PT  diagnose if male infertility or reduced sperm count has a genetic
PT  basis
PS  Claim 6; Fig 4A-B; 54pp; English.
CC  This represents the amino acid sequence of the short form of the TPR
CC  motif Y (TPRY) gene product. The invention relates to genes occurring on
CC  the non-recombining region of the human Y chromosome. The sequences fall
CC  into two classes: (1) X-homologous DNA which are expressed in many
CC  organs, having functional X homologues and (2) testis-specific DNA
CC  sequences, Y chromosomal DNA from males with known conditions such as
CC  infertility and reduced sperm count can be assessed using the invention
CC  to determine whether the condition is associated with or caused by the
CC  occurrence of the gene or gene alteration. Candidate inhibitors of the
CC  enzymatic activity of the genes can be assessed using in vitro assays.
SQ  Sequence 1079 AA;
```

alignment_scores:
```
        Quality:   280.00              Length:     85
        Ratio:       3.836             Gaps:        1
Percent Similarity: 85.882      Percent Identity: 72.941
```

alignment_block:
US-09-030-606-174/rev x W81504

Align seg 1/1 to: W81504 from: 1 to: 1079

```
1414 GCTGGAGTATAGTGGTGTGATCTCAACTCACTCAACCTCTGCCTCCCAT  1365
     ||||||:::   |||||||||:::||||||:||||||::||||||
996  AlaGlyMetGlnTrpCysAspLeuuSerSerLeuGlnProProProGl  1012

1364 ATTCAAGCAATTCCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGC  1315
     ||||||::|||||  ||||||||||||||:::::::||||||
1012 yPheLysArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArgH  1029

1314 GCCTGGCCACCATATCCAGCTAACTTTGTATTTTAGTACAGACAGGATT  1265
     ||||||:||||  :||||||||||||:||||||::||||||
1029 isLeuProSerCysProThrAsnPheCysIlePhe.ValGluThrGlyPh  1045

1264 TCACCATTTGGCCAGGCTGCTGTCTTGAACTCCTTACCTCAAGTGATCTGC  1215
     |||||| ||||||||||||||||||:::||||||
1045 eHisHisValGlyGlnAlaCysLeuGluLeuThrSerGlyGlyLeuL  1062
```

```
1214  CTGCCTC.GCCTCCCAAAGTGCTGGGATTACAGGCATGAGGCCACCTGCC 1166
      |||  |||||||||||||||||||||||||||::::||||||||  |||
1062  euAlaSerAlaSerGlnSerArgAlaGlyIleThrGlyValSerHisHisAla 1078

1165  CAG 1163
      :::
1079  Arg 1079


seq_name: A_Geneseq_36:W88627

seq_documentation_block:
ID   W88627 standard; Protein; 87 AA.
AC   W88627;
DT   01-MAR-1999 (first entry)
DE   Secreted protein encoded by gene 94 clone HPMBQ32.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9854963-A2.
PD   10-DEC-1998.
PF   04-JUN-1998; U11422.
PR   18-DEC-1997; US-070923.
PR   06-JUN-1997; US-048877.          PR   05-SEP-1997; US-057762.
PR   06-JUN-1997; US-048881.          PR   05-SEP-1997; US-057765.
PR   06-JUN-1997; US-048884.          PR   05-SEP-1997; US-057771.
PR   06-JUN-1997; US-048893.          PR   05-SEP-1997; US-057776.
PR   06-JUN-1997; US-048896.          PR   06-JUN-1997; US-048876.
PR   06-JUN-1997; US-048899.          PR   06-JUN-1997; US-048880.
PR   06-JUN-1997; US-048915.          PR   06-JUN-1997; US-048883.
PR   06-JUN-1997; US-048949.          PR   06-JUN-1997; US-048892.
PR   06-JUN-1997; US-048964.          PR   06-JUN-1997; US-048895.
PR   06-JUN-1997; US-048972.          PR   06-JUN-1997; US-048898.
PR   06-JUN-1997; US-049020.          PR   06-JUN-1997; US-048901.
PR   06-JUN-1997; US-049375.          PR   06-JUN-1997; US-048917.
PR   05-SEP-1997; US-057628.          PR   06-JUN-1997; US-048963.
PR   05-SEP-1997; US-057635.          PR   06-JUN-1997; US-048971.
PR   05-SEP-1997; US-057644.          PR   06-JUN-1997; US-049019.
PR   05-SEP-1997; US-057647.          PR   06-JUN-1997; US-049374.
PR   05-SEP-1997; US-057650.          PR   05-SEP-1997; US-057627.
PR   05-SEP-1997; US-057661.          PR   05-SEP-1997; US-057634.
PR   05-SEP-1997; US-057667.          PR   05-SEP-1997; US-057643.
PR   05-SEP-1997; US-057761.          PR   05-SEP-1997; US-057646.
PR   05-SEP-1997; US-057764.          PR   05-SEP-1997; US-057649.
PR   05-SEP-1997; US-057770.          PR   05-SEP-1997; US-057654.
PR   05-SEP-1997; US-057775.          PR   05-SEP-1997; US-057666.
PR   05-SEP-1997; US-057778.          PR   05-SEP-1997; US-057760.
PR   06-JUN-1997; US-048875.          PR   05-SEP-1997; US-057763.
PR   06-JUN-1997; US-048878.          PR   05-SEP-1997; US-057769.
PR   06-JUN-1997; US-048882.          PR   05-SEP-1997; US-057774.
PR   06-JUN-1997; US-048885.          PR   05-SEP-1997; US-057777.
PR   06-JUN-1997; US-048894.
PR   06-JUN-1997; US-048897.          PA   (HUMA-) HUMAN GENOME SCI INC.
PR   06-JUN-1997; US-048900.          PI   Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA,
PR   06-JUN-1997; US-048916.          PI   Fan P, Feng P, Ferrie AM, Fischer CL, Florence C,
PR   06-JUN-1997; US-048962.          PI   Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW,
PR   06-JUN-1997; US-048970.          PI   Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM,
PR   06-JUN-1997; US-048974.          PI   Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
PR   06-JUN-1997; US-049373.          DR   WPI; 99-059865/05.
PR   05-SEP-1997; US-057584.          DR   N-PSDB; V84504.
PR   05-SEP-1997; US-057629.          PT   New isolated human genes and the secreted polypeptides they encode -
PR   05-SEP-1997; US-057642.          PT   useful for diagnosis and treatment of e.g. cancers, neurological
PR   05-SEP-1997; US-057645.          PT   disorders, immune diseases, inflammation or blood disorders
PR   05-SEP-1997; US-057648.          PS   Claim 11; Page 530-531; 772pp; English.
PR   05-SEP-1997; US-057651.          CC   The invention relates to nucleic acid sequences (V84411 to V84633)
PR   05-SEP-1997; US-057662.          CC   encoding human secreted proteins (W88534 to W88756). The secreted protein
PR   05-SEP-1997; US-057668.          CC   gene sequences are deposited with the ATCC under deposit numbers ATCC
                                      CC   97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
                                      CC   209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
                                      CC   cells comprising recombinant vectors containing the nucleic acid
                                      CC   sequences are used for the recombinant production of the secreted
                                      CC   proteins. The polynucleotide and amino acid sequences are useful for are
                                      CC   useful for preventing, treating or ameliorating medical conditions e.g.
                                      CC   by protein or gene therapy. Pathological conditions can be also diagnosed
                                      CC   by determining the amount of the new polypeptides in a sample or by
                                      CC   determining the presence of mutations in the new polynucleotides.
                                      CC   Specific uses are described for each of the polynucleotides, based on
                                      CC   which tissues they are most highly expressed in, and include developing
                                      CC   products for the diagnosis or treatment of cancer, neurodegenerative
                                      CC   disorders, developmental abnormalities and foetal deficiencies, blood
                                      CC   diseases, tumours, leukemias, diseases of the immune system, autoimmune
                                      CC   diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
                                      CC   ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
                                      CC   restenosis, prostate diseases, obesity, disorders involving osteoclasts
                                      CC   such as osteoporosis, arthritis or malignancies, diseases of testes, lung
                                      CC   or thymus, digestive/endocrine disorders, infections and AIDS. The
                                      CC   polypeptides are also useful for identifying their binding partners.
                                      CC   The present sequence represents human secreted protein (see descriptor
                                      CC   line for gene number and clone identification).
                                      SQ   Sequence   87 AA;


alignment_scores:
            Quality:   263.50          Length:       79
            Ratio:       4.183         Gaps:          2
Percent Similarity:     79.747    Percent Identity:   69.620
```

alignment_block:
US-09-030-606-174/rev x W88627

Align seg 1/1 to: W88627 from: 1 to: 87

```
1456 TTTTTTTTTTTTTTGAGACAGAGTCTTACTCTGTTGCCCCAGCTGGAGT 1407
     ||| ||| |||    :::||| ||||||:::  ||||||||
   9 PheCysPheValPheGluMetAspSerSerValThrGlnAlaGlyVal Ar 25

1406 ATAGTGGTGGTGATCATCAACTCACTTCAACCTCTGCCTCCCATATTCAAGC 1357
     | |||||||||||| :||||||||    |||||| |||||:::
  25 lGlnTrpCysAspLeuGlySerLeuGlnAlaProProGlyPheSerP 42

1356 AATTCCTCGCCTCCAGCCTCCCAAGTAGCTGGATTACAGGCGCCTGCCA 1307
     |||||||||||||||||||||||||||||||||||||||||| |||
  42 roPheSerCysLeuSerLeuProSerSerTrpAspTyrArgArgProPro 58

1306 CCATATCCAACTTTTGTA...TTTTAGTACAGACAGGATTCACC 1260
     |||  |||||||| |||::    ||||||||||||:::||||||||
  59 ProArgProAlaAsn.PheLeuTyrPheLeuValGluThrGlyPheHisH 75

1259 ATTTTGGCCAGGCTGGTCTGTCTGAACTCCTTACCTCA 1225
     ||  ::::||  |||||||||||||||||||||
  75 isValSerGlnAspGlyLeuAspLeuThrSer 86
```

seq_name: A_Geneseq_36:W94493

seq_documentation_block:
ID   W94493 standard; Protein; 268 AA.
AC   W94493;
DT   23-APR-1999 (first entry)
DE   Human kallikrein.
KW   Human; kallikrein; keratinocyte; HKALL; skin disorder; cancer; eczema;
KW   psoriasis; scleroderma; adenocarcinoma; leukaemia; melanoma.
OS   Homo sapiens.
PN   WO9842849-A1.
PD   01-OCT-1998.
PF   25-MAR-1998; U05939.
PR   26-MAR-1997; US-824874.
PA   (INCY-) INCYTE PHARM INC.
PI   Hillman JL, Lal P;
DR   WPI; 99-070073/06.
DR   N-PSDB; X16295.
PT   Human kallikrein polypeptide, HKALL - useful e.g. to treat skin
PT   disorders e.g. eczema, psoriasis and melanoma. Antibodies specific for HKALL
PT   to treat skin disorders and cancers
PS   Claim 1; Page 41-42; 61pp; English.
CC   The present sequence represents human kallikrein, designated HKALL.
CC   HKALL (or fragments) can be administered therapeutically to increase
CC   proteolysis and subsequent skin scaling to treat/prevent skin disorders,
CC   e.g. eczema, psoriasis and scleroderma. HKALL has chemical and
CC   structural homology with human stratum corneum chymotryptic enzyme
CC   (thought to be involved in the proteolysis of intercellular cohesive
CC   structures necessary for desquamation, the process by which outer layers
CC   of skin are eliminated), and its expression in cDNA libraries was
CC   associated with tumour-associated tissues and skin cells. It can be used
CC   to screen for antagonists and agonists, and to generate antibodies.
CC   HKALL antagonists can be used to suppress excessive proteolysis and
CC   subsequent skin cell scaling, so may be administered to treat skin
CC   disorders. They may also be used to suppress excessive cell
CC   proliferation, so can be administered to treat/prevent cancer, e.g.
CC   adenocarcinoma, leukaemia and melanoma. Antibodies specific for HKALL
CC   may be used directly as antagonists, or indirectly as a targeting or
CC   delivery mechanism for bringing pharmaceutical agents to HKALL-expressing
CC   cells. They are also useful to diagnose conditions/diseases characterised
CC   by HKALL expression and to monitor therapeutic interventions. The
CC   polynucleotide encoding HKALL, or complementary sequences, can be used to
CC   produce hybridisation probes, useful to detect polynucleotides encoding
CC   HKALL, e.g. to diagnose diseases relating to polypeptide expression
CC   (e.g. cancers of the bladder, prostate) or monitor HKALL regulation
CC   during therapeutic intervention. Polynucleotides encoding HKALL are
CC   useful to produce antisense sequences for therapeutic administration to
CC   modulate/prevent HKALL expression e.g. to treat/prevent skin disorders

CC   or cancer a s above.
SQ   Sequence   268 AA;

alignment_scores:
     Quality: 256.00   Length: 173
     Ratio: 2.560   Gaps: 4
     Percent Similarity: 57.803 Percent Identity: 35.838

alignment_block:
US-09-030-606-174 x W94493

Align seg 1/1 to: W94493 from: 1 to: 268

```
   3 TCAGCGCCACACTGTTTCCAGAAGTGAGTGCAGAGTCCTACACCATCGG 52
     :::||||||||||||| ||||||||||||||       ::: :::|||
  80 ThrAlaAlaHisCysArgLysLys..........ValPheArgValAr 92

  53 GCTGGGCCTGCACAGTCTTGAGGCCGACCAGGAGCCAGGAGCCAGATGG 102
     ||||||   |||:||||::::::  |||||||:::||||| |||:||||
  92 gLeuGlyHisTyrSerLeuSerProValTyrGluSerGlyGlnMetP 109

 103 TGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTC 152
     ::::::: :::::: |||||||  :::||||
 109 heGlnGlyValLysSerIleProHisProGlyTyrSerHisProGlyHis 125

 153 GCTAAGCACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGA 202
     :::||||||||||||||||||||||||:::
 126 SerAsnAspLeuMetLeuIleLysLeuAsnArgArgileArgProThrLy 142

 203 CACCATCCGGAGCATCAGCATTGCTTCGCAGTGCCTACCGCGGGGAACT 252
     |||||| ||||||::||:::|||||:::||||||||||||||:::||||
 142 sAspValArgProIleAsnValSerSerHisCysProSerAlaGlyThrL 159

 253 CTTGCCCGTTCTGGGTGGGTCGTCTGCTGGCGAACGTGAGCTCACGGGT 302
     ::||||||||||||||||||
 159 ysCysLeuValSerGlyTrpGly............................ 166

 303 GTGTGTCTGCCCTCTTCAAGGAGGTCCTCTGCCCAGTCGCGGGGGCTGAC 352
                                                     |  ||
 167 ................................................Th 167

 353 CCAGAGCCTCGCGTCCCAGGCAGCAGAATGCCTACCGTGCTGCAGTGCGTGAA 402
     |    :::||| ||:::  |||| |||||||||||||||||||||:::||
 167 rThrLysSerProGlnVal..HisPheProLysValLeuGlnCysLeuAs 183

 403 CGTGTCGTGGTGTCTGANGAGGTCTGCANTAAGCTCTATGACCCGCTGT 452
     |:::||:::||| :::  ||        ||   ||   ||:::|||||
 183 nIleSerValLeuSerGlnLysArgCysGluAspAlaTyrProArgGlnI 200

 453 ACCACCCCANCATGTTCTGCCCGCCGGGAGGGCAAGACCAGAAGGACTCC 502
     ::||||||||||||  || :::  :::||||||
 200 leAspAspThrMetPheCysAla...GlyAspLysAlaGlyArgAspSer 215

 503 TGCAAGCTGAGAGAGGGG 520
     |||::: |||||||
 216 CysGlnGlyAspSerGly 221
```

seq_name: A_Geneseq_36:Y02693

seq_documentation_block:
ID   Y02693 standard; Protein; 110 AA.
AC   Y02693;
DT   11-JUN-1999 (first entry)
DE   Human secreted protein encoded by gene 44 clone HTDAD22.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

```
OS   Homo sapiens.
PN   WO9902546-A1.
PD   21-JAN-1999.
PF   07-JUL-1998;  U13684.
PR   12-SEP-1997;  US-058785.
PR   08-JUL-1997;  US-051916.
PR   08-JUL-1997;  US-051918.
PR   08-JUL-1997;  US-051919.
PR   08-JUL-1997;  US-051920.
PR   08-JUL-1997;  US-051925.
PR   08-JUL-1997;  US-051926.
PR   08-JUL-1997;  US-051928.
PR   08-JUL-1997;  US-051929.
PR   08-JUL-1997;  US-051930.
PR   08-JUL-1997;  US-051931.
PR   08-JUL-1997;  US-051932.
PR   08-JUL-1997;  US-052732.
PR   08-JUL-1997;  US-052733.
PR   08-JUL-1997;  US-052793.
PR   08-JUL-1997;  US-052795.
PR   08-JUL-1997;  US-052803.
PR   18-AUG-1997;  US-055684.
PR   18-AUG-1997;  US-055722.
PR   18-AUG-1997;  US-055723.
PR   18-AUG-1997;  US-055947.
PR   18-AUG-1997;  US-055948.
PR   18-AUG-1997;  US-055949.
PR   18-AUG-1997;  US-055950.
PR   18-AUG-1997;  US-055953.
PR   18-AUG-1997;  US-055954.
PR   18-AUG-1997;  US-055964.
PR   18-AUG-1997;  US-055984.
PR   18-AUG-1997;  US-056360.
PR   18-AUG-1997;  US-058660.
PR   12-SEP-1997;  US-058661.
PR   12-SEP-1997;  US-058664.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DW, Li Y, Moore PA,
PI   Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;
DR   WPI: 99-120770/10.
DR   N-PSDB: X27354.
PT   New isolated human genes and the secreted polypeptides they encode -
PT   useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 11; Page 343; 464pp; English.
CC   This sequence represents a secreted human protein encoded by the gene
CC   clone detailed in the descriptor line.  The gene can be used to generate
CC   fusion proteins by linking to the gene to a human immunoglobulin Fc
CC   portion (e.g. X27302) for increasing the stability of the fused protein
CC   as compared to the human protein only.
CC   The invention relates to 123 novel genes and their fragments (nucleic
CC   acid sequences: X27311-X27449; amino acid sequences Y02650-Y02788) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 123 polynucleotides, based on
CC   which tissues they are most highly expressed in (see X27311 for described
CC   uses).
SQ   Sequence   110 AA;

alignment_scores:
          Quality:   249.00            Length:       93
          Ratio:      3.662            Gaps:          3
Percent Similarity:  73.118      Percent Identity:  60.215

alignment_block:
US-09-030-606-174/rev x Y02693

Align seg 1/1 to: Y02693 from: 1 to: 110

.1455 TTTTTTTTTTTTTTGAGACAGAGTCTTACTCTGTTGCCCAGCTGGAGTA 1406
```

```
         |||   |||||||||||||   |||||||||||||::::|||
    21 PheValPheLeuArgLeuSerLeuLeuSerLeuLeuProLysLeu..... 35

  1405 TAGTGGTGTGATCTCAACCTTCAACCCTGCCTCCCATATTCAAGCA 1356
         |||  :::||:::||:::||:::|||||||||||::::||:|::::|||
    36 .Glu..CysAsnLeuGlySerLeuGlnProProProArgPheGlnAr 51

  1355 ATTCTCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCTGCCAC 1306
         :||||||||||||||||||  :::||||||||||||||||| |||
    51 gPheSerCysLeuSerLeuLeuAsnSerTrpAspTyrArgArgProProP 68

  1305 CATATCCAGCTAACTTTTGTATTTTTAGTACAGACAGGATTCACCATTT 1256
         ||:::  |||||||||::: |||  |||||||||||||  ::||||
    68 roHisLeuAlaAsnPheCysValSerArgGlyGlyValSerSerCys 84

  1255 TGGCCAGGCTGGTCTTGAACCTCCTTACCTCAAGTGACTGCCTGCCTCGC 1206
         |||||||||||||||||||| ||||| ||||
    85 TrpProGlyTrpSerArgThrProAspLeuMetIleArgLeu.ProArgP 101

  1205 CTCCCAAAGTGCTGGGATTACAGGCA 1180
         ||||||:::||||||||||||||||
   101 roProArgValLeuGlyLeuGlnAla 109

seq_name: A_Geneseq_36:W48351

seq_documentation_block:
ID   W48351 standard; Protein; 123 AA.
AC   W48351;
DT   14-AUG-1998 (first entry)
DE   Human breast cancer related protein BCRB2.
KW   L-oncogene; diagnosis; treatment; ovarian cancer; antibody; antisense.
OS   Homo sapiens.
FH   Key               Location/Qualifiers
FT   Binding_site      5. .10
FT                     /note= "haem iron binding site"
FT   Domain            26. .46
FT                     /note= "Transmembrane helix""
FT   Domain            79. .108
FT                     /note= "Transmembrane helix"
PN   WO9807851-A2.
PD   26-FEB-1998.
PF   22-AUG-1997;  E04600.
PR   22-AUG-1996;  CA-183900.
PA   (BERG/) BERGMANN J E.
PA   (PRED/) PREDDIE E R.
PI   Bergmann JE, Preddie ER;
DR   WPI: 98-169156/15.
DR   N-PSDB: V17747.
PT   Human breast cancer related genes - used for diagnosis,
PT   pre-symptomatic detection and therapy of breast and ovarian cancers
PS   Claim 2; Fig 3e; 96pp; English.
CC   The proteins encoded by human breast cancer related genes, L-oncogenes
CC   (V17738-V17753) can be used in the diagnosis and treatment of breast and
CC   ovarian cancer.  Antibodies and multiple antigenic peptide epitopes can
CC   be used to detect the presence of the proteins.  The antibodies can also
CC   be used to block the activity of the proteins.  Antisense molecules can be
CC   used to prevent expression of the proteins.
SQ   Sequence   123 AA;

alignment_scores:
          Quality:   249.00            Length:       80
          Ratio:      4.150            Gaps:          1
Percent Similarity:  75.000      Percent Identity:  63.750

alignment_block:
US-09-030-606-174/rev x W48351

Align seg 1/1 to: W48351 from: 1 to: 123

  1398 GTGATCTCAACTCACTTCAACCTCTGCCTCCCATATTCAAGCAATTCTCC 1349
         :::||||::::|||  ::: |||  |||||||||||||||
     1 MetIleThrAlaHisCysLysHisCysHisProSerSerAsnSerPr 17
```

```
1348 TGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCGCCCTGCCACCATATCC 1299
     ||||||    :::|||||:::    |||||||    ||:::|||
  17 oAlaSerProThrGlnValValGluIleThrGlyIleCysTyrHisAlaT 34

1298 AGCTAACTTTGTATTTTAGTACAGACAGGATTTCACCATTTGGCCAG 1249
     |||  |||||||||||||:::    |||||||||||   |||||
  34 rpLeuIlePheValGluMetGlyPheHisHisValGlyGln 50

1248 GCTGGTCTTGAACTCCTTACCTCAAGTGATCTGCCTGCCTCGCCTCCCAA 1199
     |||||||||||||||||||||||| :::||||     |||
  51 AlaGlyLeuGluLeuLeuThrSer.MetIleThrCysLeuGlyHisProL 67

1198 AGTGCTGGGATTACAGGCATGAGCCACCCTGCCCAGCC 1161
     ||||||||||||||||    |||||||    |||||
  67 ysCysTrpAspTyrArgPheGluProProArgProAla 79
```

This Page Blank (uspto)

```
****************************************************

    |\  /| |¯¯| |¯¯| |¯¯| |¯¯| |¯¯|
    | \/ | |__| |__| |__| |  | |__|
    |    | |    |    |  \ |  | |  |  (TM)

****************************************************

    Release 3.1A John F. Collins, Biocomputing Research Unit.
    Copyright (c) 1993-1998 University of Edinburgh, U.K.
        Distribution rights by Oxford Molecular Ltd

****************************************************

MPsrch_tpn   n.a. - n.a. Smith-Waterman search, using a-protein query
             which has been backtranslated into n.a. using IUPAC symbols

Run on:        Wed Sep 29 01:43:37 1999;    MasPar time 913.24 Seconds
                                        1447.804 Million cell updates/sec

Tabular output not generated.

Title:          >US-09-030-606-172
Description:    (1-159) from US09030606.pep
Perfect Score:  1871
N.A. Sequence:  1 ATGGTNGARGCNWSNSNYTNWS...........ARAARACNGTNCARGCNWSN 477
Comp:                TACCANCTYCGNWSNRANWS...........TYTYTGNCANGTYCGNWSN

Scoring table:  TABLE bktranslate2
                Gap 30

Nmatch  STD :   Dbase 0;  Query 0

Searched:       646147 seqs, 1385953633 bases x 2

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:       embl58
                1:em_bal  2:em_ba2  3:em_fun  4:em_htg  5:em_hum1  6:em_hum2
                7:em_in   8:em_om   9:em_or  10:em_ov  11:em_pat 12:em_ph
               13:em_pl  14:em_ro  15:em_sts 16:em_vi

Database:       genbankill
               17:gb_ba1 18:gb_ba2 19:gb_htg1 20:gb_htg2 21:gb_in1
               22:gb_in2 23:gb_om 24:gb_ov 25:gb_pat 26:gb_ph 27:gb_pl1
               28:gb_pl2 29:gb_prl 30:gb_pr2 31:gb_pr3 32:gb_ro
               33:gb_st 34:gb_sts 35:gb_sy 36:gb_un 37:gb_vi

Statistics:     Mean 61.585;  Variance 203.336;  scale 0.303

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                          SUMMARIES
                 %
Result  Query
No.  Score Match Length DB  ID          Description                Pred. No.
-----------------------------------------------------------------------------
 1   1789  95.6  1347  31  AF113140   Homo sapiens serine pr      5.01e-197
 2   1379  73.7  1140  23  SSU76256   Sus scrofa enamel matr      4.35e-145
 3   1234  66.0  1237  32  AF019979   Mus musculus enamel ma      7.27e-127
 4    704  37.6  5900  31  AF113141   Homo sapiens serine pr      1.95e-61
 5    489  26.1   732  25  I95869     Sequence 2 from patent      6.21e-36
 6    489  26.1  1419  29  D78203     Homo sapiens mRNA for       6.21e-36
 7    489  26.1  1438  25  E13202     Human gene for serine       6.21e-36
 8    489  26.1  1451  30  AF013988   Homo sapiens serine pr      6.21e-36
 9    489  26.1  1506  30  HSU62801   Human protease M mRNA,      6.21e-36
10    477  25.5   969  31  HUMSERPROT Human stratum corneum       1.52e-34
11    477  25.5   986  25  A42048     Sequence 1 from Patent      1.52e-34
12    420  22.4   881  32  RNTRYPVB   Rat mRNA for trypsin V      5.16e-28
```

```
c 13  406  21.7  586  34  G41906     SHGC-56840 Human Homo      1.98e-26
  14  400  21.4  877  32  RNTRYPVA   Rat mRNA for trypsin V     9.40e-26
  15  365  19.5  428  32  MMTAM1     Mouse mRNA for gamma-7     7.62e-22
  16  357  19.1  777  32  S48142     T-kininogenase-kallikr     5.83e-21
  17  355  19.0  849  32  MMNGFB     Mouse mRNA for gamma s     9.69e-21
  18  355  19.0  860  24  GGU15157   Gallus gallus trypsino     9.69e-21
  19  352  18.8  786  32  AB016032   Mus musculus mRNA for      2.07e-20
  20  352  18.8  858  32  MMPRCE     Mouse mRNA for a prore     2.07e-20
  21  352  18.8 1110  32  AF016269   Rattus norvegicus myel     2.07e-20
  22  347  18.5  870  32  MUSEGFBPM  Mouse major epidermal      7.34e-20
  23  346  18.5  893  24  PPTRYP     P.platessa mRNA for tr     9.45e-20
  24  347  18.5 1256  14  AB016227   Mus musculus mRNA for      7.34e-20
  25  342  18.3  731  24  AF089847   Deinagkistrodon acutus     2.59e-19
  26  341  18.2 1091  32  AB015206   Mus musculus mRNA for      3.33e-19
  27  341  18.2 1122  32  MMU18723   Mus musculus mRNA for      3.33e-19
  28  336  18.0 1570  24  TSU21903   Trimeresurus stejneger     1.17e-18
  29  333  17.8  814  32  MMTRYAR    Mouse mRNA for preprot     2.48e-18
  30  333  17.8  879  24  DMU58945   Dissostichus mawsoni p     2.48e-18
  31  332  17.7  500  32  MUSGKAL    Mouse glandular kallik     3.19e-18
  32  332  17.7  850  32  MMPRECE2   Mouse mRNA for a prore     3.19e-18
  33  332  17.7  869  32  MUSEGFBPB  Mouse epidermal growth     3.19e-18
  34  332  17.7 1186  29  AB012917   Homo sapiens mRNA for      3.19e-18
  35  329  17.6  907  24  AF012463   Pleuronectes americanu     6.76e-18
  36  326  17.4  833  24  AF015727   Gloydius halys salmobi     1.43e-17
  37  326  17.4 1390  24  AF056033   Gloydius halys salmobi     1.43e-17
  38  326  17.4 1555  24  AF018568   Gloydius ussuriensis c     1.43e-17
  39  326  17.4 1563  24  D67083     Trimeresurus gramineus     1.43e-17
  40  322  17.2  500  32  MMKA01     Mouse mRNA encoding ka     3.87e-17
  41  321  17.2  877  24  PMTRYPSIN  P.magellanica mRNA for     4.96e-17
  42  321  17.2  978  24  AF012462   Pleuronectes americanu     4.96e-17
  43  317  16.9  518  32  RATKALLSA  Rat submaxillary gland     1.34e-16
  44  317  16.9  723  32  RATKALLA   Rattus norvegicus (clo     1.34e-16
  45  317  16.9  762  32  RATKALLB   Rattus norvegicus (clo     1.34e-16
```

                                  ALIGNMENTS

```
RESULT     1
LOCUS      AF113140   1347 bp   mRNA      PRI      25-MAR-1999
DEFINITION Homo sapiens serine protease prostase mRNA, complete cds.
ACCESSION  AF113140
NID        g4512029
VERSION    AF113140.1   GI:4512029
KEYWORDS   .
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1347)
AUTHORS    Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
           Wang,K.
TITLE      Molecular cloning and characterization of prostase, an androgen
           regulated serine-protease with prostate-restricted expression
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
REFERENCE  2 (bases 1 to 1347)
AUTHORS    Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
           Wang,K.
TITLE      Direct Submission
JOURNAL    Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
           Bothell, WA 98021, USA
FEATURES             Location/Qualifiers
     source          1..1347
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19q13"
     CDS             1..765
                     /note="androgen regulated"
                     /codon_start=1
                     /product="serine protease prostase"
                     /protein_id="AAD21580.1"
                     /db_xref="PID:g4512030"
                     /db_xref="GI:4512030"
```

```
                    /translation="MATAGNPWGWFLGYLILGVAGSLVSGSCSQIINGEDCSPHSQPW
                    QAALFLEDDFFCGGVLVHPQWVLSAAHCFQNSYTIGLGLHLEPEQEPGSQMVEFASLS
                    VRHPEYNRPLLANDLMLIKLDESVESDTIRSISIASQCPTPAGNSCLVSGWGLLANGR
                    MPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGHDQKDSCNGDSGGPLICNGYLQ
                    GLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS"
  polyA_site        1327..1332
BASE COUNT      269 a      489 c      334 g      255 t
ORIGIN

Query Match                95.6%;    Score 1789;  DB 31;  Length 1347;
Best Local Similarity   55.5%;    Pred. No. 5.01e-197;
Matches  264;  Conservative  114;  Mismatches  98;  Indels    0;  Gaps    0;

Db     286  ATGGTGGAGGCAGCCTCTCCCTACGGCACCCAGAGTACACAGACCCTGCTCGCTAAC 345
            ||||| |||:||| :|:: || :| |||:| ||||:| ||:| :| ||| ||| ||
Qy       1  ATGGTNGARGCNWSNYTNWSNGTNMNGCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY  60

Db     346  GACCTCATGCTCATCAAGTTGGACAGAATCCGTGTCCGAGTCTGACCACCATCCGGAGCATC 405
            |||:| |||:|| ||||:| ||:|| ||:|| ::|: || || ||||  :::||
Qy      61  GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db     406  AGCATTGCTTCGCAGTGCCCTACCCGCGGGAACTCTTGCCTGTTTCGGCTGGGGTCTG 465
            :: ||:||| .:|: ||||:| || |||||| || :|:||| || | |||| :|| |
Qy     121  WSNATHGCNWSNCARTGYCCNACNCGNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db     466  CTGGCAACGGCAGAATGCCTACCGTGCTGCTGCAGTGCGTGAACGTGTCGGTGGTGTCTGAG 525
            ::| ||:|| |   |||||: | |:|| ||:| ||||:| |||:| ||:||| |:|| :||
Qy     181  YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db     526  GAGGTCTGCAGTAAGCTCTATGACCCCAGCCTGTACCACCCAGCATGTTTCTGCCGCGGCCGA 585
            |||:|:|| ::| |::||:| |: ::  ||| || ||::|| ||::|::|| ||:||||
Qy     241  GARGTNTGYWSNAARYTNTAYGAYCCNYNTAYCAYCCNWSNATGTTYTGYCCNGGNGGN 300

Db     586  GGGCACGACCAGAAGGACTCCTGCAACGGTGAACCGCCCCTGATCTGCAACGGG 645
            || | || ||:|: ||:|| | |||:| |  ::| ||:|| ||:|| ||:
Qy     301  GGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGN 360

Db     646  TACTTGCAGGGCCTTGTGTCTTTCGGAAAAGCCCCGTGTGGCCAGGTGCCAGGT 705
            || |:||| ||:||||| || ::|| || |||||:|| ||||:| |||| || ||
Qy     361  TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNCCNGGN 420

Db     706  GTCTACACCAACCTCTGCAAATTCACTGAGTGGATAGAGAAAACCGTCAGGCCAG 761
            ||:||:|| |::| | ||:|: ||:||:|| |||||||| | ||| :| |:|
Qy     421  GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAAARACNGTNCARGCNWS 476
```

```
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     sig_peptide     69..140
                     /note="the signal peptide is cleaved after Ala24 and the
                     preprotein is secreted into the developing enamel matrix"
     CDS             69..833
                     /note="EMSP1"
                     /codon_start=1
                     /product="enamel matrix serine proteinase 1 precursor"
                     /protein_id="AAB94638.1"
                     /db_xref="PID:g2737921"
                     /db_xref="GI:2737921"
                     /translation="MTAAGNPWGWFLGYLILLGITGSLAFINGGHIINGEDCNPHSQPW
                     QAALFLEDDFFCGGVLVHPQWVLSAAHCFQNSYTIGLGLHNLEPEQEPGSQMMEASLS
                     IQHPEYNEPSMANDLMLIKLKESVSLSDTVRNISVVSQCPTPGDSCLVSGWGRLASGR
                     LPQVLQCVNISVASEEVCKARYGPVYHPSMFCAGGQQDKDSCHGDSGGPLICNGSLQ
                     GLVSFGQAQCGQPNVPGVYTNLCKFTDWIQTTIQAS"
     mat_peptide     159..830
                     /note="the preprotein is cleaved following His30
                     activating the serine proteinase; the active protein has
                     an apparent molecular weight of 34 kDa and a derived
                     molecular weight of 24 kDa"
                     /product="enamel matrix serine proteinase 1"
     misc_feature    order(177..179,549..551)
                     /note="disulfide bond"
     misc_feature    order(234..236,282..284)
                     /note="disulfide bond"
     misc_feature    join(279..281,414..416,687..689)
                     /note="encodes catalytic triad"
     misc_feature    405..407
                     /note="encodes potential glycosylation location; yields a
                     blank cycle during protein sequencing"
     misc_feature    order(408..410,789..791)
                     /note="disulfide bond"
     misc_feature    order(510..512,705..707)
                     /note="disulfide bond"
     misc_feature    order(675..677,750..752)
                     /note="disulfide bond"
BASE COUNT      245 a      366 c      295 g      234 t
ORIGIN

Query Match                73.7%;    Score 1379;  DB 23;  Length 1140;
Best Local Similarity   48.1%;    Pred. No. 4.35e-145;
Matches  229;  Conservative  108;  Mismatches  139;  Indels    0;  Gaps    0;

Db     354  ATGATGGAGGCCAGCCTCTCCATTCAGCACCCAGAATACAACAGCAGCCGTCTATGGCCAAC 413
            |||| ||||:|| :|:: :| :| |||:| :||:|::||||||:|:||||::|| :|||
Qy       1  ATGGTNGARGCNWSNYTNWSNGTNMNGCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY  60

Db     414  GACCTCATGCTCATCAAGTTGAAAGAATCGGTGCGTCGCTGCTCTGACCACCGTCCGGAACATC 473
            |||:| |||:|| ||||:| ||:|| ||:|| ::|: || || ||||  :::||
Qy      61  GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db     474  AGCGTCGTCTCCCAGTGCCGACCCCTGGGATTCTTGCCTCGTTCCGGCTGGGGTCGG 533
            || :|:|| :|:|| |::|: :| |:|||||: :| |:: || |:||| |:| :||
Qy     121  WSNATHGCNWSNCARTGYCCNACNCGNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db     534  CTGGCCAGTGCAGACTGCAGACTGCCCCAAGTGTCGCAGTGTGAACATCTCCGTGGCCTTCTGAG 593
            ::| ||:|| | |||| ||:| :: ||:| ||:|| | | || |||| ::| || :|| :||
Qy     181  YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db     594  GAGGTCTGCAAGGCGGCGCTACGGCCCCGTGTACCACCCCAGACATGTTCTGCGCGTGGCCGA 653
            |||:|:|| || ||:|  |: || ::  ||| || ||::|| ||::|::|| ||:||||
Qy     241  GARGTNTGYWSNAARYTNTAYGAYCCNYNTAYCAYCCNWSNATGTTYTGYCCNGGNGGN 300

Db     654  GGCCAGGACCAAAAGGACTTCCTGCCACGTGACTCTGGGGCCCCTGATCTGCAATGGG 713
            || | || ||:|: ||:|:|| | :|||:| || |:||| ||:|| ||:|| ||:
Qy     301  GGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGN 360

Db     714  TCCCTGCAGGGCCTCGTGCTCTTTGGACAAGCCCAGTGTGGCCAACCAACGTGCCAGGC 773
            | ::||| ||:||||| || ::|| || |||||:|| ||||:| |||| || ||:|:
Qy     361  TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 420
```

Left column:

```
Db  774  GTCTACCACCAACCTCTGCAAATTCACGGACTGGATACAGAGCGACCATCCAGGCCAG 829
         || ||:|| ||:|| || |||:||:||:|| ||||| || :|| |||||| ::
Qy  421  GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476

RESULT   3
LOCUS     AF019979      1237 bp    mRNA          ROD       01-JAN-1999
DEFINITION Mus musculus enamel matrix serine proteinase 1 precursor, mRNA,
          complete cds.
ACCESSION AF019979
NID       g4090846
VERSION   AF019979.1  GI:4090846
KEYWORDS  .
SOURCE    house mouse.
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Euteleostomi; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1  (bases 1 to 1237)
AUTHORS   Simmer,J.
TITLE     Enamel Matrix Serine Proteinase 1 (EMSP1)
JOURNAL   Unpublished
REFERENCE 2  (bases 1 to 1237)
AUTHORS   Simmer,J.
TITLE     Direct Submission
JOURNAL   Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas
          Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
          Antonio, TX 78284-7888, USA
FEATURES        Location/Qualifiers
     source     1..1237
                /organism="Mus musculus"
                /strain="Swiss-Webster"
                /db_xref="taxon:10090"
                /cell_type="ameloblast"
     sig_peptide 52..126
     CDS        52..819
                /function="degradation of the enamel matrix during enamel
                maturation"
                /note="EMSP1"
                /codon_start=1
                /product="enamel matrix serine proteinase 1 precursor"
                /protein_id="AAC98894.1"
                /db_xref="PID:g4090847"
                /db_xref="GI:4090847"
                /translation="MMVTARTPWGWFLGCLILEVTGASASASVSSRIIQGQDCSPHSQP
                WQAALFSEDGFFCSGVLVHPQWVLSAAHCLQESYIVGLGLHNLKGSQEPGSRMLEAHL
                SIQHPNFNDPSFANDLMLJKLNESVIESNTIRSIPVATOCPTPGDTCLVSGWGQLKNG
                KLPSLLQCVNLSVASEETCRLLIYDPVYHLSMFCAGGGQDOKDSCNGDSGGPIVCNRSL
                QGLVSMGQGKCGQPGIPSYVTNLCKFTNWIQTIIQTN"
     misc_feature 127..144
                /note="encodes propeptide"
     mat_peptide 145..816
                /product="matrix serine proteinase 1"
BASE COUNT   289 a   386 c   386 g   276 t
ORIGIN
```

```
Query Match   66.08%;  Score 1234;  DB 32;  Length 1237;
Best Local Similarity 46.38%;  Pred. No. 7.27e-127;
Matches  220;  Conservative 100;  Mismatches 155;  Indels  0;  Gaps   0;

Db  340  ATGTTAGAGGCCCACCTCTCCATCCAGCACCCCAACTTCAACGATCCTTCTTTTGCAAAC 399
         || ||| ||| | ::| |||| || || || |||:|| : || |||| :| | |||||
Qy    1  ATGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAYAMGNCCNYTNYTNGCNAAY 60

Db  400  GATCTCATGCTCATCAAACTGAAGGAGTCAGTGATGATAGAGTCTAACACTATCAGGAGCATC 459
         ||:: |||:|||| ||| |: ::|  || | :|:|| ||:| |||| |: ||| |:||| ||
Qy   61  GAYYTNATGYTNATHAARYTNGAYCAYGARWSNGTNWSNGARWSNSGAYCNATHMGNWSNATH 120

Db  460  CCTGTGGCTACCCAATGCCCCACTCCTGGAGATACCTGCCTAGTCTCTGGTTGGGGTCAA 519
         : |||| ||::|||| : || ||| ||:|| ||:| ||: |||:| || || |||||| :
Qy  121  WSNATHGCNWSNCARTGYCCNACNCCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db  520  CTAAAGAATGGAAACTGCCCAGCCTCCTGCAGTGTGTGAATCCTGCAGTGGCGTCAGAG 579
         :|  |||| | |:||| ||  |  || ||| | || | ||| ||:|| |: || : ||
```

Right column:

```
Qy  181  YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db  580  GAGACTTGCCGCTGCTGTGATGACCCTGTGTACCACCTCAGTAGTTCTGCCTGGAGGA 639
         || :|| || :|| ||:||||| | |  ||||| :: ||||||| |:|| ||:::|
Qy  241  GARGTNTGYWSNAARYTNTAYGAYCCNYTNPAYCAYCCNWSNATGTTYTGYCNGCNGGNGGN 300

Db  640  GGACAGGACCAAAAGGACTCCTGCAATGGGGACTCTGGGGTCCCATAGTGTGCAACAGA 699
         ||||: |||: ||:|| ||  ||| | || | ||||| || |  ::||||| | |||:|
Qy  301  GGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGN 360

Db  700  TCCCTACAAGGCTCGTCTGCTCTATGGGACAAGGAAAGTGTGGGCAGCCTGGCATACCAAGT 759
         |  || ||:||   |:| || || ||:|  |:||||||:|| |  |:||  || | |:|||
Qy  361  TAYYTNCARGGNYTNGTNWSNTTYGGNAARGNCNCCNTGYGGNCARGGNGTNCCNGGN 420

Db  760  GTCTACCACCAATCTCTGCCAAGTTCACTAACTGGATCCAGACCATCATTCAGACAA 814
         ||:||:|| |||||| ||:||||||  ||:|||| | |  |:| :|| :|||| ||
Qy  421  GTNTAYCACNAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNW 475

RESULT   4
LOCUS     AF113141      5900 bp    DNA           PRI       25-MAR-1999
DEFINITION Homo sapiens serine protease prostase gene, complete cds.
ACCESSION AF113141
NID       g4512031
VERSION   AF113141.1  GI:4512031
KEYWORDS  .
SOURCE    human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 5900)
AUTHORS   Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
          Wang,K.
TITLE     Molecular cloning and characterization of prostase, an androgen
          regulated serine-protease with prostate-restricted expression
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
REFERENCE 2  (bases 1 to 5900)
AUTHORS   Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
          Wang,K.
TITLE     Direct Submission
JOURNAL   Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
          Bothell, WA 98021, USA
FEATURES        Location/Qualifiers
     source     1..5900
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="19"
                /map="19q13"
     mRNA       join(<1411..1471,2735..2897,3319..3569,3653..3789,
                5062..5817)
                /product="serine protease prostase"
     exon       1411..1471
                /number=1
     CDS        join(1411..1471,2735..2897,3319..3569,3653..3789,
                5062..5214)
                /note="androgen regulated"
                /codon_start=1
                /product="serine protease prostase"
                /protein_id="AAD21581.1"
                /db_xref="PID:g4512032"
                /db_xref="GI:4512032"
                /translation="MATAGNPWGWFLGYLLIGVAGSLVSGGSCSQIINGEDCSPHSQPW
                QAALVMENELFCSGVLVHPQWVLSAAHCFQNSYTIGLGLHSLEADQEPGSOMVEASLS
                VRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGR
                MPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGHDQKDSCNGDSGGPLICNGYLQ
                GLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS"
     exon       2735..2897
                /number=2
     exon       3319..3569
                /number=3
     exon       3653..3789
                /number=4
     exon       5062..5817
```

```
BASE COUNT        1417 a      1569 g      1271 t
                  /number=5
ORIGIN

Query Match                 37.6%;   DB 31;  Length 5900;
Best Local Similarity       51.8%;   Pred. No. 1.95e-61;
Matches         99;  Conservative     53;  Mismatches   39;  Indels    0;  Gaps    0;

Db   3380  ATGTGAGGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTCGCTAAC 3439
            ||||| ||||||| :::: :|| :| |||:||||::|:||:||| :| || || ||:||
Qy      1  ATGTGNARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY   60

Db   3440  GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATC 3499
            |||:|||| |||||||| :|| :: ||:||| ||| || |:||: || :: :|| |||
Qy     61  GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH  120

Db   3500  AGCATTGCTTCGCGAGTGCCCTACCGCGGGGAACTCTTGCCTGTTCTTGGCTGGGTCTG  3559
            :: |:::: :: ||::|||| :: ||||| ||| ||:|||:||||:|||||| :|
Qy    121  WSNATHGCNWSNCARTGYCCNACNCGNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTN  180

Db   3560  CTGGCGAACGG 3570
            :| ||:|:||:|
Qy    181  YTNGCNAAYGG  191

RESULT    5
LOCUS     I95869          732 bp    DNA                        PAT    17-JUL-1998
DEFINITION  Sequence 2 from patent US 5733768.
ACCESSION I95869
NID       g3940339
VERSION   I95869.1  GI:3940339
KEYWORDS  Unknown.
SOURCE    Unknown.
ORGANISM  Unclassified.
REFERENCE 1  (bases 1 to 732)
AUTHORS   Dixon,E.P., Johnstone,E.M., Little,S.P. and Norris,F.H.
TITLE     Amyloid precursor protein protease
JOURNAL   Patent: US 5733768-A 2 31-MAR-1998;
FEATURES             Location/Qualifiers
source               1..732
                     /organism="unknown"

BASE COUNT         168 a      212 c      203 g      149 t
ORIGIN

Query Match                 26.1%;   DB 25;  Length 732;
Best Local Similarity       36.2%;   Pred. No. 6.21e-36;
Matches        165;  Conservative     71;  Mismatches  220;  Indels    0;  Gaps    0;

Db    275  CTGTGATCCACCCTGACTATGATGACCCCAGCCATGACCAGGACATCATGCTGTTCGCGCC  334
            ::| || ||| ||||:|| ||: |:  :|||||:||| || :|||||||:| ||||:|
Qy     20  SNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARY   79

Db    335  TGGCACGCCCCAGCCAAACTCTCTGAACTCATCCAGCCCCTTCCCCTGGAGAGGGACTGCT  394
            |||| ||||| :: ||| | :: || |||| ||:|| |:  ||:::| || ||| ||
Qy     80  TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNATGGCNWSNCARTGYC  139

Db    395  CAGCCAACAACCAGCTGCCACATCCTGGGCTGGGCAAGACAGCAGATGGTGATTCC  454
            ||| |:| |: :: ||| :: |:||| |||  :: |:| |: |:| :|||  |
Qy    140  CNACNCGNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTNGCNAAYGGNGNATGC  199

Db    455  CTGACACCATCCAGTGTGCATACATCCCACCTGGTGTCCCGTGAGGAGTGTCAGCATGCCT  514
            ||:|| ||||| |||::||:: :: || :||||| |:||| || || ||:|| ||
Qy    200  CNACNGTNYTNCARTGYGTNAAYGTNTNGTNWSNGARGARGTNTGTWSNAARYTNT  259

Db    515  ACCCTGGCCAGATCACCCAGAACATGTTGTGTCGGTCTGGGGATGAGAAGTACGGGAAGGATT  574
            | :: ||| ||| |: :: ||:: ::| |:| ||  |:|| || :||| :::|| || ||
Qy    260  AYGAYCCNYTNTAYCAYCCNWSNARTGTTYTGYCNGNGNGNGNCARNNNCARNNNGAYW  319

Db    575  CCTGCCAGGTGATTCTGGGGTCCGTGGTATGTGAGAGCACCTCCGAGGCCCTTGTGT  634
            | :|| ||| | ::|| |||| :| ||| |:|| |:: || |:| ||:|| |:|
Qy    320  SNTGYAAYGGNGAYWSNGGNGGNGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNW  379
```

FEATURES   Location/Qualifiers
    source   1..1438
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"

BASE COUNT   334 a   432 c   362 g   310 t
ORIGIN

Query Match   26.1%;   Score 489;   DB 25;   Length 1438;
Best Local Similarity   36.2%;   Pred. No. 6.21e-36;
Matches   165;   Conservative   71;   Mismatches 220;   Indels   0;   Gaps   0;

Db   430   CTGTGATCCACCTGACTATGATGCCGCCAGCCATGACCAGGACATCATGCTGTTGCGCC   489
           | : :|| : ||:|:|| ||:|: : |: : : |:: | |:|
Qy   20   SNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARY   79

Db   490   TGGCACGCCCAGCCAAACTCTCTGAACTCATCCAGCCCCTTCCCCTGGAGAGGGACTGCT   549
           | : : |||| |:|:|| | |:: | :| |:| :|
Qy   80   TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNATHGCNWSNCARTGYC   139

Db   550   CAGCCAACACCACCAGCTGCCACATCCTGGGCTGGGCAAGACAGCAGATGGTGATTTCC   609
           |-| | ::: ||:: || | : |||
Qy   140   CNACNGCNGGNAAYWSNGTYYTNGTNWSNGGNYTNYTNGCNAAYGGNMGNATGC   199

Db   610   CTGACACCATCCAGTGTGCATACATCCACCTGGTGTCCCGTGAGGAGTGTGAGCATGCCT   669
           | |:::| ||::: | |: ::| |||| :|
Qy   200   CNACNGTNYTNCARTGYGTNAAYGTNWSNGTNTNWSNGARGARGTNTGYWSNAARYTNT   259

Db   670   ACCCTGGCCAGATCACCCAGAACATGTTGTGTGCTGGGGATGAGAAGTACGGGAAGGATT   729
           |: : : :| : || |: || | | |:: | :|
Qy   260   AYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGNGNGGNCARNNNGAYW   319

Db   730   CCTGCCAGGTGATTCTGGGGTCCGCTGGTATGTGGAGACCACCTCCAGGCCTTGTGT   789
           : ||:| ||:: || |||: |: : :| ||| : :|| :|
Qy   320   SNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNW   379

Db   790   CATGGGGTAACATCCCCTGTGATCAAAGGAGAAGCCAGGAGTCTACACCAACGTCTCA   849
           : : || | |: || :| | || :|: ||:|:|
Qy   380   SNTTYGGNAARGCNCCNTGYGGNCARGTNGGNTNCCNGGNGTNTAYACNAAYYTNTGYA   439

Db   850   GATACACGAACTGGATCCAAAAAACCATTCAGGCCA   885
           :|:|| ||||: ||:::|| | |:|| :
Qy   440   ARTTYACNGARTGGATHGARARACNGTNCARGCNW   475

RESULT   7

LOCUS   E13202   1438 bp   DNA   PAT   24-JUN-1998
DEFINITION   Human gene for serine protease,complete cds.
ACCESSION   E13202
NID   g3252007
VERSION   E13202.1   GI:3252007
KEYWORDS   JP 1997149790-A/3.
SOURCE   Homo sapiens.
  ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1438)
  AUTHORS   Tsuruoka,N., Yamashiro,K., Tsujimoto,M. and Yamaguchi,M. .
  TITLE   NEW SERINE PROTEASE
  JOURNAL   Patent: JP 1997149790-A 3 10-JUN-1997;
            SUNTORY LTD
COMMENT   OS   Homo sapiens (human)
           PN   JP 1997149790-A/3
           PD   10-JUN-1997
           PF   24-JUL-1996 JP 1996212196
           PR   29-SEP-1995 JP 95P 275105
           PI   TSURUOKA NOBUO, YAMASHIRO KYOKO, TSUJIMOTO MASAFUMI, PI
           YAMAGUCHI MARE
           PC   C12N15/09,C07H21/04,C07K14/47,C12N1/21,C12N5/10, PC
           C12N9/52//A61K38/46,
           PC   (C12N15/09,C12R1:91),(C12N1/21,C12R1:19),(C12N9/52,C12R1:19),
           PC   (C12N9/52,
           PC   C12R1:91);
           CC   strandedness: Double;
           CC   topology: Linear;
           FH   Key   Location/Qualifiers
           FH
           FT   source   1..1438
           FT   /organism='Homo sapiens'
           FT   /cell_type='colon cancer-derived cell' FT
                /cell_line='COLO 201'
                /clone='SP59'
           FT   CDS   156..890
           FT   /product='Serine protease precursor' FT
           FT   sig_peptide   156..218
           FT   mat_peptide   219..887
           FT   /product='serine protease'.

Qy   20   SNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARY   79

Db   490   TGGCACGCCCAGCCAAACTCTCTGAACTCATCCAGCCCCTTCCCCTGGAGAGGGACTGCT   549
Qy   80   TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNATHGCNWSNCARTGYC   139

Db   550   CAGCCAACACCACCAGCTGCCACATCCTGGGCTGGGCAAGACAGCAGATGGTGATTTCC   609
Qy   140   CNACNGCNGGNAAYWSNGTYYTNGTNWSNGGNYTNYTNGCNAAYGGNMGNATGC   199

Db   610   CTGACACCATCCAGTGTGCATACATCCACCTGGTGTCCCGTGAGGAGTGTGAGCATGCCT   669
Qy   200   CNACNGTNYTNCARTGYGTNAAYGTNWSNGTNTNWSNGARGARGTNTGYWSNAARYTNT   259

Db   670   ACCCTGGCCAGATCACCCAGAACATGTTGTGTGCTGGGGATGAGAAGTACGGGAAGGATT   729
Qy   260   AYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGNGNGGNCARNNNGAYW   319

Db   730   CCTGCCAGGTGATTCTGGGGGTCCGCTGGTATGTGGAGACCACCTCCAGGCCTTGTGT   789
Qy   320   SNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNW   379

Db   790   CATGGGGTAACATCCCCTGTGATCAAAGGAGAAGCCAGGAGTCTACACCAACGTCTCA   849
Qy   380   SNTTYGGNAARGCNCCNTGYGGNCARGTNGGNTNCCNGGNGTNTAYACNAAYYTNTGYA   439

Db   850   GATACACGAACTGGATCCAAAAAACCATTCAGGCCA   885
Qy   440   ARTTYACNGARTGGATHGARARACNGTNCARGCNW   475

RESULT   8

LOCUS   AF013988   1451 bp   mRNA   PRI   12-AUG-1997
DEFINITION   Homo sapiens serine protease mRNA, complete cds.
ACCESSION   AF013988
NID   g2318114
VERSION   AF013988.1   GI:2318114
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;
            Homo.
REFERENCE   1 (bases 1 to 1451)
  AUTHORS   Little,S.P., Johnstone,E.M., Dixon,E.P., Norris,F., Buckley,W.,
            Becker,G., Johnson,M.,., Dobbins,J.R., Wyrick,T., Miller,J.R.,
            MacKellar,W., Hepburn,D., Corvalan,J., McClure,D., Liu,X.,
            Stephenson,D. and Clemens,J.
  TITLE   Zyme cDNA isolated from AD brain tissue
  JOURNAL   J. Biol. Chem. (1997) In press
REFERENCE   2 (bases 1 to 1451)
  AUTHORS   Little,S.P., Johnstone,E.M. and Norris,F.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JUL-1997) CNS Division, Eli Lilly and Company, Lilly
            Corporate Center, Indianapolis, IN 46285, USA
FEATURES   Location/Qualifiers
    source   1..1451
                 /organism="Homo sapiens"
                 /strain="human"
                 /db_xref="taxon:9606"

REFERENCE   2 (bases 1 to 1506)
AUTHORS   Anisowicz,A., Sotiropoulou,G. and Sager,R.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUL-1996) Cancer Genetics, Dana-Farber Cancer
             Institute/Harvard Medical School, 44 Binney, Boston, MA 02115, USA
FEATURES   Location/Qualifiers
     source   1..1506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="76N normal mammary epithelial cells"
                     /chromosome="19"
                     /map="19q13.3"
     sig_peptide   246..293
     CDS   246..980
                     /note="serine protease homolog of the trypsin family"
                     /codon_start=1
                     /product="protease M"
                     /protein_id="AAB07113.1"
                     /db_xref="PID:g1518788"
                     /db_xref="GI:1518788"
                     /translation="MKKLMVVLSLIAAAWAEEQNKLVHGGPCDKTSHPYQAALYTSGH
                     LLCGGVLIHPLWVLTAAHCKKPNLQVFLGKHNLRQRESSQEQSSVVRAVIHPDYDAAS
                     HDQDIMLLRLARPAKLSELIQPLPLERDCSANTTSCHILGWGKTADGFPDTIQCAYI
                     HLVSREECEHAYPGQITQNMLCAGDEKYGKDSCQGDSGGPLVCGDHLRGLVSWGNIPC
                     GSKEKPGVYTNVCRYTNWIQKTIQAK"
     misc_feature   294..308
                     /note="encodes activation peptide"
     mat_peptide   309..977
                     /product="protease M"
     misc_feature   429..431
                     /note="encodes His-62 catalytic triad"
     misc_feature   561..563
                     /note="encodes Asp-106 catalytic triad"
     misc_structure 645..653
                     /note="encodes glycosylation motif: AsnThrThr"
     misc_feature   816..818
                     /note="encodes active site for trysin-like cleavage:
                     Asp-191"
     misc_feature   834..836
                     /note="encodes Ser-197 catalytic triad"
     polyA_signal   1095..1100
                     /note="pancreas-specific"
     polyA_signal   1490..1495
BASE COUNT   321 a   465 c   396 g   324 t
ORIGIN

Query Match   26.1%;   Score 489;   DB 30;   Length 1506;
Best Local Similarity   36.2%;   Pred. No. 6.21e-36;
Matches   165; Conservative   71; Mismatches 220;   Indels   0;   Gaps   0;

Db   520 CTGTGATCCACCCTGACTATGATGCGCGCCAGCCATGACCAGGACATCATGCTGTTGCGCC 579
                : ||| |||||| || || | |:: ||| |:| :||:|:| | ||||:|:
Qy   20 SNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARY   79

Db   580 TGGCACGCCCAGCCAAACTCTCTGAACTCATCCAGCCCCTTCCCGTGGAGAGGGACTGCT 639
                 | |||| ::||| |||| :: || ||| ||| | ||| || :| || :| |
Qy   80 TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHGCNWSNCARTGYC 139

Db   640 CAGCCAACACCACCAGCTGCCACATCCTGGGCTGGGGCAAGACAGCAGATGGTGATTTCCC 699
                  | || | | :||||| |||||||||||| ||| |||: |||| |||||
Qy   140 CNACNGCNGGNAAYWSNTGYTNGGGNTNGGGNTGGGGNTNYTNGCNAAYGGNMGNATGC 199

Db   700 CTGACACGAACTGGATCCAAAAAAACCATTCAGGCCA 759
             ::| || |||||| :::|| | || ||||||||
Qy   200 CNACNGTNTYNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNT 259

Db   760 ACCCTGCCAGATCACCCCAGAACATGTTGTGCTGGGATGAGAAGTACGGAAGGATT 819
                | :: :| |||||||| ||| | :: :| :|| :||| :||: |:||
Qy   260 AYGAYCCNYTNTAYCAYCCNWSNATGTTYTYGYCNGGNGNGGNGGNCARNNNGAYW 319

Db   820 CCTGCCAGGTGATTCTGGGGGTCCGCTGGTATGTGGAGACCACCTCCAGGCCTTGTGT 879
             : |||:|| ||: :| || | | :| : :: || || : | :: :: ||
Qy   320 SNTGYAAYGGNGAYWSNGGNGNCCNYTNATHTGYAAYGGNTAYYTNGTNGTNW 379

RESULT   9
LOCUS   HSU62801   1506 bp   mRNA   PRI   02-SEP-1996
DEFINITION Human protease M mRNA, complete cds.
ACCESSION   U62801
NID   g1518787
VERSION   U62801.1   GI:1518787
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1506)
AUTHORS   Anisowicz,A., Sotiropoulou,G., Stenman,G., Mok,S.C. and Sager,R.
TITLE   A Novel Protease Homolog Differentially Expressed in Breast and
            Ovarian Cancer
JOURNAL   Mol. Med. (Camb. Mass.) (1996) In press

CDS
                     /chromosome="19"
                     /map="19q13.3"
                     /tissue_type="Alzheimer's disease brain tissue"
            147..881
                     /note="zyme; protease bears homology to Kallikrein class
                     and can be localized to microvessels and microglia;
                     chymotrypsin-like"
                     /codon_start=1
                     /product="serine protease"
                     /protein_id="AAB66483.1"
                     /db_xref="PID:g2318115"
                     /db_xref="GI:2318115"
                     /translation="MKKLMVVLSLIAAAWAEEQNKLVHGGPCDKTSHPYQAALYTSGH
                     LLCGGVLIHPLWVLTAAHCKKPNLQVFLGKHNLRQRESSQEQSSVVRAVIHPDYDAAS
                     HDQDIMLLRLARPAKLSELIQPLPLERDCSANTTSCHILGWGKTADGFPDTIQCAYI
                     HLVSREECEHAYPGQITQNMLCAGDEKYGKDSCQGDSGGPLVCGDHLRGLVSWGNIPC
                     GSKEKPGVYTNVCRYTNWIQKTIQAK"
BASE COUNT   342 a   434 c   367 g   308 t
ORIGIN

Query Match   26.1%;   Score 489;   DB 30;   Length 1451;
Best Local Similarity   36.2%;   Pred. No. 6.21e-36;
Matches   165; Conservative   71; Mismatches 220;   Indels   0;   Gaps   0;

Db   421 CTGTGATCCACCCTGACTATGATGCGCGCCAGCCATGACCAGGACATCATGCTGTTGCGCC 480
                : ||| |||||| || || | |:: ||| |:| :||:|:| | ||||:|:
Qy   20 SNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARY   79

Db   481 TGGCACGCCCAGCCAAACTCTCTGAACTCATCCAGCCCCTTCCCGTGGAGAGGGACTGCT 540
                 | |||| ::||| |||| :: || ||| ||| | ||| || :| || :| |
Qy   80 TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHGCNWSNCARTGYC 139

Db   541 CAGCCAACACCACCAGCTGCCACATCCTGGGCTGGGGCAAGACAGCAGATGGTGATTTCC 600
                  | || | | :||||| |||||||||||| ||| |||: |||| |||||
Qy   140 CNACNGCNGGNAAYWSNTGYTNGGGNTNGGGNTGGGGNTNYTNGCNAAYGGNMGNATGC 199

Db   601 CTGACACCATCCAGTGCATACATCCTGGTGTCCCGTGAGAGTGTGAGCATGCCT 660
             ::| || | |:|| :: || ||| || || || || || :| |:| |:||
Qy   200 CNACNGTNTYNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNT 259

Db   661 ACCCTGCCAGATCACCCCAGAACATGTTGTGCTGGGGATGAGAAGTACGGGAAGGATT 720
             | :: :| |||||||| ||| | :: :| :|| :|| :||: |:||
Qy   260 AYGAYCCNYTNTAYCAYCCNWSNATGTTYTYGYCNGGNGNGGNGGNCARNNNGAYW 319

Db   721 CCTGCCAGGTGATTCTGGGGGTCCGCTGTATGTGGAGACCACCTCCAGGCCTTGTGT 780
             : |||:|| ||: :| || | | :| : :: || || : | :: :: ||
Qy   320 SNTGYAAYGGNGAYWSNGGNGNCCNYTNATHTGYAAYGGNTAYYTNGTNGTNW 379

Db   781 CATGGGGTAACATCCCCGTGGATCAAAGGAGAAGCCAGGAGTCTACACCACCGTCTCGCA 840
             : ::|| ::|| | || |:|| | || || | || || |:|| ||: :| :|
Qy   380 SNTTYGGNAARGCNCCNTGYGGNCARGTNGGNTNCCNGGNGTNTAYACNAAYYTNTGYA 439

Db   841 GATACACGAACTGGATCCAAAAAAAACCATTCAGGCCA 876
              :| || | ||||| ::::|| | || ||||||||
Qy   440 ARTTYACNGARTGGATHGARAARACNGTNCARGCNW 475

```
Db   880  CATGGGGTAACATCCCTGTGGATCAAAGGAGGAGGCCAGGAGTCTCACACCACGTCTGCA  939
               : ||:||  ||  |||||| |||:||  |||:|||  ||  |||| ||  ||| ||  ||:|
Qy   380  SNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGNGNTNTAYACNAAYTNTGYA   439

Db   940  GATACACGAACTGGATCCAAAAAACCATTCAGGCCA  975
               :|: ||:||||||  ||||||||||||  |:|||:|
Qy   440  ARTTYACNGARTGGATHGARAARACNGTNCARGCNW  475
```

RESULT   10
LOCUS   HUMSERPROT   969 bp   mRNA   PRI   03-DEC-1998
DEFINITION   Human stratum corneum chymotryptic enzyme mRNA, complete cds.
ACCESSION   L33404
NID   g521214
VERSION   L33404.1 GI:521214
KEYWORDS   serine proteinase.
SOURCE   Homo sapiens cDNA to mRNA.
  ORGANISM   Homo sapiens
             Eukaryota; mitochondrial eukaryotes; Metazoa; Chordata;
             Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 969)
  AUTHORS   Hansson,L., Stromqvist,M., Backman,A., Wallbrandt,P., Carlstein,A.
             and Egelrud,T.
  TITLE   Cloning, expression, and characterization of stratum corneum
             chymotryptic enzyme. A skin-specific human serine proteinase
  JOURNAL   J. Biol. Chem. 269 (30), 19420-19426 (1994)
  MEDLINE   94308225
FEATURES   Location/Qualifiers
     source   1..969
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="keratinocyte"
     CDS   16..777
                     /note="SCCE; serine proteinase"
                     /codon_start=1
                     /product="stratum corneum chymotryptic enzyme"
                     /protein_id="AAC37551.1"
                     /db_xref="PID:g532504"
                     /db_xref="GI:532504"
                     /translation="MARSLLLPLQILLLSLALETAGEEAQGDKIIDGAPCARGSHPWQ
                     VALLSGNQLHCGGVLVNERWVLTAAHCMNEYTVHLGSDTLGDRRAQRIKASKSFRHP
                     GYSTQTHVNDLMLVKLNSQARLSSMVKKVRLPSRCEPPGTTCTVSGWGTTTSPDVTFP
                     SDLMCVDVKLISPQDCTKVYKDLLENSMLCAGIPDSKKNACNGDSGGPLVCRGTLQGL
                     VSWGTFPCGQPNDPGVYTQVCKFTKWINDTMKKHR"
BASE COUNT   256 a   281 c   233 g   199 t
ORIGIN

Query Match   25.5%; DB 31; Length 969;
Best Local Similarity   36.8%; Pred. No. 1.52e-34;
Matches 173; Conservative   84; Mismatches 207; Indels 207; Gaps   4;

```
Db   296  AGGCCTCGAAGTCATTCCGCCACCCGCCACTCCACACAGACCCATGTAATGACCTCA  355
               |:||   :: :  |:||:||  :|:||||||| |||||||:|:||:|
Qy     8  ARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNA  67

Db   356  TGCTCGTGAAGCTCAATAGCCAGGCCAGGCTGTCATCCATGGTGAAGAAAGTCAGGCTGC  415
               ||| |:  :||||:|  ::||| || : :|:|| :| |||:|
Qy    68  TGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGTNWSNGAYACNATHMGNWSNATHG  127

Db   416  CCTCCCGCTGCGAACCCCCTGTACTGTCTCCGGCCGGCCACTACCACGA  475
               |||:||| |:: ||| :| |:|||||| |||||| ||:
Qy   128  CNWSNCARTGYCCNACNGCNGGNAAYWSNGTYYTNGGNGNGGGN-YTNYTN--  183

Db   476  GCCCAGATGTGACCTTCCCTGACCTCATGTGCGTGGATGTCAAGCTCATCTCCCCC  535
               || :|| : |:  ||  ::|| |||:||| |||||||
Qy   184  GCN-A-AYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARG  241

Db   536  AGGACTGCCACGAAGGTTACAAGGACTTACTGGAAAATTCCATGCTGTGCGCCTGGCATCC  595
               ||: ||::|||  ::  |||  ::|| :|||||| |:|| ||
Qy   242  ARGTNTGYWSNAARYTNTAYGAYCCNTNTAYCAYCCNWSNATGTTYTGYGCNGNGNGNG  301

Db   596  CCGACTCCAAGAAAAACGCCTCAATGGTGACCGCAGGGGACCGTTGGTGTGCAGAGTA  655
```

Qy 302 GNCARNNNCARNNNGAYWSNGTYAAYGNGGNGAYWSNGGNGNCNYTNATHTGYAAYGGNT 361
```

---

Qy   302  GNCARNNNCARNNNGAYWSNGTYAAYGNGGNGAYWSNGGNGNCNYTNATHTGYAAYGGNT  361
               |:   .:||:||:||  :|||:|:  ||  ||||  |
Db   656  CCCTGCAAGGTCGTGGTGCCTGGGGAACTTCCCTGCGGGCCAACCCAATGACCCAGGAG  715
               ::|||:||||  ||| |  |:||| :|  ||||:||:
Qy   362  AAYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNG  421

Db   716  TCTCACACTCCAAGTGCTGCCAAGTTCACCAAGTGGATAAATGACACCATGAAA  765
               : |||:|| ||| :  ||||||||||  |:|||||| |:||||| :
Qy   422  TNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCAR  471

RESULT   11
LOCUS        A42048                986 bp    DNA             PAT       05-MAR-1997
DEFINITION   Sequence 1 from Patent WO9500651.
ACCESSION    A42048
NID          g2297542
VERSION      A42048.1  GI:2297542
KEYWORDS
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 986)
  AUTHORS    Egelrud,T. and Hansson,L.
  TITLE      RECOMBINANT STRATUM CORNEUM CHYMOTRYPTIC ENZYME (SCCE)
  JOURNAL    Patent: WO 9500651-A 1 05-JAN-1995;
             SYMBICOM AB (SE)
COMMENT      Other publication CA 2165197 950105
             Other publication AU 6935294 950117
             Other publication PL 312189 960401
             Other publication FI 956075 960216
             Other publication NO 955110 960215.
FEATURES             Location/Qualifiers
     source          1..986
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     sig_peptide     25..90
     CDS             25..786
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAA02644.1"
                     /db_xref="PID:e305178"
                     /db_xref="PID:g2297543"
                     /db_xref="GI:2297543"
                     /translation="MARSLLLPLQILLLSLALETAGEEAQGDKIIDGAPCARGSHPWQ
                     VALLSGNQLHCGGVLVNERWVLTAAHCMNEYTVHLGSDTLGDRRAQRIKASKSFRHP
                     GYSTQTHVNDLMLVKLNSQARLSSMVKKVRLPSRCEPPGTTCTVSGWGTTTSPDVTFP
                     SDLMCVDVKLISPQDCTKVYKDLLENSMLCAGIPDSKKNACNGDSGGPLVCRGTLQGL
                     VSWGTFPCGQPNDPGVYTQVCKFTKWINDTMKKHR"
BASE COUNT      260 a     286 c     237 g     203 t
ORIGIN

Query Match             25.5%;  DB 25;  Length 986;
Best Local Similarity   36.8%;  Pred. No. 1.52e-34;
Matches  173;  Conservative   84;  Mismatches 207;  Indels 207;  Gaps   4;

```
Db   305  AGGCCTCGAAGTCATTCCGCCACCCGCTACTCCACACAGACCCATGTAATGACCTCA  364
               |:||   .:: :  :||:||  :|:|||| |||||||:|:||:|
Qy     8  ARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNA  67

Db   365  TGCTCGTGAAGCTCAATAGCCAGGCCAGGCTGTCATCCATGGTGAAGAAAGTCAGGCTGC  424
               |:||  |:  :||||:|  ::||| || :  :|:|| :| |||:|
Qy    68  TGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNATHG  127

Db   425  CCTCCCGCTGCGAACCCCCTGGAACCACCTGACTGTCTCCGGCCACTACCACGA  484
               |||:||| |::  ||| :| |:|||||| |||||| ||:
Qy   128  CNWSNCARTGYCCNACNGCNGGNAAYWSNGTYYTNGGNGNGGGN-YTNYTN--  183

Db   485  GCCCAGATGTGACCTTCCCTGACCTCATGTGCGTGGATGTCAAGCTCATCTCCCCC  544
               || :|| :  |:  ||  ::|| |||:||| |||||||
Qy   184  GCN-A-AYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGNARG  241
```

```
Db   545  AGGACTGCACGAAGGTTTACAAGGACTTACTGGAAAAATTCCATGCTGCGGCTGGCATCC  604
          |:| ||::|| ||: |:|   |   ||   ::: ||| |  :: || ||::| ||
Qy   242  ARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCCNWSNATGTTYTGYCNGNGGNGNG     301

Db   605  CCGACTCCAAGAAAAACGCCTGCAATGTGACTGGGGGGACCGTTGGTGTGCAGAGGTA    664
          :|: ||::||| ||::|| || ||: || ||| |  | ||| ||
Qy   302  GNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGNCCNYTNATHTGYAAYGGNT  361

Db   665  CCCTGCCAAGGTCTGGTGTCCTGGGGAACTTCCCTTGCGGCCAACCCAATGACCCAGGAG  724
          ::| ||:||  |  || | ||  ||  |:|| |  || :|||| |||||: ||  || ||
Qy   362  AYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNG  421

Db   725  TCTACACCTCAAGTGTGCGAAGTTCACCAGGTGGATAAATGACACCATGAAA          774
          | | |::|| || || || |  |  | ||||  ||   |:: |    || |
Qy   422  TNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAAARACNGTNCAR           471


RESULT      12
LOCUS       RNTRYPVB        881 bp    mRNA          ROD      17-MAR-1992
DEFINITION  Rat mRNA for trypsin v b-form.
ACCESSION   X59013
NID         g57414
VERSION     X59013.1  GI:57414
KEYWORDS    trypsin.
SOURCE      black rat.
ORGANISM    Rattus rattus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 881)
AUTHORS     Kang,J.
TITLE       Direct Submission
JOURNAL     Submitted (12-APR-1991) J. Kang, Institut f Genetik, Universitaet
            zu Koeln, Weyertal 121, D-5000 Koeln, FRG
REFERENCE   2 (bases 1 to 881)
AUTHORS     Kang,J., Wiegand,U. and Muller-Hill,B.
TITLE       Identification of cDNAs encoding two novel rat pancreatic serine
            proteases
JOURNAL     Gene 110 (2), 181-187 (1992)
MEDLINE     92165057
FEATURES             Location/Qualifiers
     source          1..881
                     /organism="Rattus rattus"
                     /db_xref="taxon:10117"
                     /dev_stage="adult"
                     /tissue_type="pancreas"
     mRNA            1..881
                     /note="trypsin V b-form"
                     /evidence=experimental
     CDS             35..775
                     /codon_start=1
                     /product="trypsin V b-form"
                     /protein_id="CAA41752.1"
                     /db_xref="PID:g57415"
                     /db_xref="GI:57415"
                     /db_xref="SWISS-PROT:P32822"
                     /translation="MKICIFFTLLGTVAAFPTEDNDDRIVGGYTCQEHSVPYQVSLNA
                     GSHICGGSLITDQWVLSAAHCYHPQLQVRLGEHNIYEIEGAEQFIDAAKMILHPDYDK
                     WTVDNDIMLIKLKSPATLNSKVSTIFLPYCCPTAGTECLVSGWGVLKFGFESPSVLQC
                     LDAPVLSDSVCHAYPRQITNNMFCLGFLEGGKDSCQYDSGGPVVCNGEVQGIVSWGD
                     GCALEGKPGVYTKVCNYLNWIQQTVAAN"
     sig_peptide     35..79
                     /product="trypsin V b-form"
     mat_peptide     107..772
                     /product="trypsin V b-form"
BASE COUNT      213 a    211 c    200 g    257 t
ORIGIN

Query Match       22.4%;   Score 420;  DB 32;  Length 881;
Best Local Similarity 38.2%;   Pred. No. 5.16e-28;
Matches  178;  Conservative  71;  Mismatches 211;  Indels   6;  Gaps    5;

Db    30 G__TTGATGGAGCCAAGATGATTGATTCTTCATCCTGACTATGATAAGTGGACTGTTGATAATGACA  359
```

```
Qy     5  TNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCNYTNYTNGCNAAYGAYY  64
          | ||||          :    :    |   |  |  |   |        ||  ||  |||
Db   360  TCATGCTGATTAAGTTGAAGTCACCAGCCACCCTGAAGTATCTAAAGTATCGACATCCTC  419
          | ||:| |||:|| :       || | | |||:| :| ||| |   | :: ::  ||:
Qy    65  TNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNA  124
          || || |  |:|  | : |   ||   :: |||  |::| || || |  || :|   ||
Db   420  TGCCACAGTACTGTCCGACAGCTGGTACTGACGTGCCTGGTGTCTGCTGGGGTGTCTTGA  479
          -   | ||  ||  | |:   || |||  ||   | ||  |  :| ::|||: |  ||
Qy   125  THGCNWSNCARTGYCCNACNGCNGGNAAYWSNGTYYTNGTNWSNGGNTGGGGNYTNYTNG  184
          || ||| |  | || :||  || | |  ||   | ||||  |||  | ||  || :|
Db   480  AATTGGCTTTGAGAGTCCTTCTGTTCTTCGTTCAGTGTCTGGATGCTCCAGTCCTGTCGATT  539
          ::|| | |||  | :||  ||  || | | :||   |  || |||   || | ::
Qy   185  CNAAYGGNMGN-AT-G-CCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARG  241
          ||  | |||   :  : |  ||  |  :|  |  :| |  ||  |   :|  |||  ||
Db   540  CTGTTTGTCACAAGGCTACCCACGTCAGATCACTCAGATCACTCAACAACATGTTCTGTCGGCTTCC  599
          ||  |:   ||  :    : |:|:  :  ::|   :||::||  | || |    ||| |
Qy   242  ARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYCNGNGGNGNG  301
          |   || |  | ||  ||   || | | ||   || ||   ||| || ||:  |||:| ||
Db   600  TGGAAGGTGGAAAGGACTCTTGCCGCTATGACGTCTGGTGGCCCTGGTGGTTGCAATGGAG  659
          ||  |||  ||||:| || |  |  ||||  | | |||  | ||  || |||  | ||
Qy   302  GNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGNCCNYTNATHTGYAAYGGNT  361
          |  ||   |||  ||  ||  ||||  |||  ||  | |  |:  :|  || ||  ||||
Db   660  AAGTCCAGGGTATTGTTCCTGGGGTGATG-GC--TGTGCTTTGGAAGGGAAGCCTGGTG  716
          ::| | ||  || | |  | |  | | ||       | | |||   ||  |||   | ||
Qy   362  AYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNG  421
          | | ||  ||  || || |   | ||  ||| |   |  || || |  ||  | || |
Db   717  TCTACACCAAGGTCTGCAACTACCTGAACTGGATTCAGCAGACCGT  762
          | | |:  |||| || ||||| ||  |||||| |:  |  | :|
Qy   422  TNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAAARACNGT  467


RESULT      13
LOCUS       G41906          586 bp    DNA            STS      29-SEP-1998
DEFINITION  SHGC-56840 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION   G41906
NID         g3668239
VERSION     G41906.1  GI:3668239
KEYWORDS    STS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 586)
AUTHORS     Myers,R.M.
TITLE       Human STSs (1998)
JOURNAL     Unpublished (1998)
COMMENT     Contact: Richard M. Myers
            Stanford Human Genome Center (SHGC)
            Stanford University School of Medicine
            Department of Genetics, M-344, Stanford, CA 94305, USA
            Tel: 4157259687
            Fax: 4157259689
            Email: myers@shgc.stanford.edu
            Primer A: AGAGACACGGTCAGCCCAAT
            Primer B: GCCAACTCCTGAGTCATCCC
            STS size: 188
            PCR Profile:
                 Initial incubation: 95 degrees C for 10 minutes

                 Denaturation:        94 degrees C for 30 seconds
                 Annealing:           60 degrees C for 30 seconds
                 Polymerization:      72 degrees C for 23 seconds
                 PCR Cycles:          30
                 Thermal Cycler:      Perkin Elmer 9700

            Protocol:
                 Template:            25 ng
                 Primer:              each 1 uM
                 dNTPs:               each 200 uM
                 AmpliTaq Gold Polymerase: 0.07 units/ul
                 Total Vol:           5 ul
```

```
               Buffer:    MgCl2:      2.5 mM
                          KCl:        50 mM
                          Tris-HCl:   10 mM
                          pH:         8.3

               Prepared with primer pairs derived from W73140 -- Unigene.
FEATURES       Location/Qualifiers
   source      1..586
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /map="19"
               /clone_lib="Human"
   STS         167..354
   primer_bind 167..186
   primer_bind complement(335..354)
BASE COUNT     130 a     139 c     179 g     136 t
ORIGIN

Query Match          21.7%;
Best Local Similarity  46.6%;   Score 406; DB 34;   Length 586;
Matches  81;  Conservative 31;  Mismatches 62;  Indels 0;  Gaps 0;   2 others

Db  351  TGGCCTGATGAGTTTCCTGGATCCACTTGGTGAACTTGCAGAGGTTCGTGTAGACACCCG  410
         ::|||| ||||| |||:||  ::||| |||||| :|:||||||||| ||:|| |:|
Cp  475  WNGCYTGNACNGTYTYTYCDATCCAYTCNGTRAAYTTRCANARRTTNGTRTANACNCCNG  416
Db  411  GTCTGTTGGGCGGCCACAAGGGTAATCTCCCCAGGACACGCGAGTCCCTGCAGGGAGCCAT  470
         |:|  ||| |  ||:|| ||::| |:|:||||||| |  ::|:: ||||||| |||| |
Cp  415  GNACNCCNACYTGNCCRCANGGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCCRT  356
Db  471  TGCAGACCACCAGGCCCCCAGAATCACCCTGGCAAGAGTCTCTACTGCTTTGTC  524
         ||:| |  ::||:|| |||:|| :|:||||||| |||||||||:||| |||
Cp  355  TRCADATNARNGGNCCNCNCCNSWRTCNCCRTTRCANSWRTCNNNYTGNNNYTGNC  302

RESULT  14
LOCUS       RNTRYPVA          877 bp    mRNA              ROD       17-MAR-1992
DEFINITION  Rat mRNA for trypsin V a-form.
ACCESSION   X59012
NID         g57412
VERSION     X59012.1    GI:57412
KEYWORDS    trypsin.
SOURCE      black rat.
ORGANISM    Rattus rattus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (bases 1 to 877)
AUTHORS     Kang,J.
TITLE       Direct Submission
JOURNAL     Submitted (12-APR-1991) J. Kang, Institut f Genetik, Universitaet
            zu Koeln, Weyertal 121, D-5000 Koeln, FRG
REFERENCE   2  (bases 1 to 877)
AUTHORS     Kang,J., Wiegand,U. and Muller-Hill,B.
TITLE       Identification of cDNAs encoding two novel rat pancreatic serine
            proteases
JOURNAL     Gene 110 (2), 181-187 (1992)
MEDLINE     92165057
FEATURES    Location/Qualifiers
   source   1..877
            /organism="Rattus rattus"
            /db_xref="taxon:10117"
            /dev_stage="adult"
            /tissue_type="pancreas"
   mRNA     1..877
            /note="trypsin V a-form"
            /evidence=experimental
   CDS      35..775
            /codon_start=1
            /product="trypsin V a-form"
            /protein_id="CAA41751.1"
            /db_xref="PID:g57413"
            /db_xref="GI:57413"
```

```
            /db_xref="SWISS-PROT:P32821"
            /translation="MKICIFFTLLGTVAAFPTEDNDDRIVGGYTCQEHSVPYQVSLNA
            GSHICGGSLITDQWVLSAAHCYHPQLQVRLGEHNYIEIEGAEQFIDAAKMILHPDYDK
            WTVNDIMLIKLKSPATLNSKVSTIPLPQYCPTATECLVSGWGVLKFGFESPSVLQC
            LDAPVLSDSVCHKAYPRQITNNMFCLGFLEGGKDSCQYDSGGPVVCNGEVQGIVSWGD
            GCALEGKPGVYTKVCNYLNWIHQTIAEN"
   sig_peptide  35..79
                /product="trypsin V a-form"
   mat_peptide  107..772
                /product="trypsin V a-form"
BASE COUNT   214 a     200 c     202 g     261 t
ORIGIN

Query Match          21.4%;        Score 400; DB 32;   Length 877;
Best Local Similarity  38.0%;   Pred. No. 9.40e-26;
Matches  176;  Conservative 70;  Mismatches 211;  Indels 6;  Gaps 5;

Db  300  TTGATGCAGCCAAGATGATTCTTCATCCTGACTATGATAAGTGGACTGTTGATAATGACA  359
         :|||| || : |||  |  |||:|| |:|||| |||:||| :||||||:||| ::||
Qy    5  TNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYY  64
Db  360  TCATGCTGATTAAGTTGAAGTCACCAGCCACCCTGAACTTAAAGTATCTACGATCCCTC  419
         |:||| |||| |||:|||| |:||:|| |||||||||  || |||:||| ||| ::||
Qy   65  TNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNA  124
Db  420  TGCCACAGTACTGTCCGACAGCTGGTACTGGGTCTGAGTGGCGTCTGGTGTCTGCGGTGTTCTGA  479
         | |:|| |||| ||||:||:|| ||| |||:| :||| ||::| ||||| |||:|||:|
Qy  125  THGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTNYTNG  184
Db  480  AATTTGGCTTTGAGAGTCCTTCTGTTCTGTCGTCTCAGTGTCTGATGCTCCAGTCCTGTCTGATT  539
         :| |||| || |||:|| |||| |||| |:||:| |||||| |||| |:||||| |:||::
Qy  185  CNAAYGGNMGN-AT-G-CCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARG  241
Db  540  CTGTTTGTCACAAGGCCTACCACGTCAGATCACTAACAACACATGTTCTGTCTGCGGCTTCC  599
         ||||:||| || ||  |:|| ||  ||: || : ||:||||| ||| |||:|| ||:||||
Qy  242  ARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGNG  301
Db  600  TGGAAGGTGGAAAGGACTCTTGCCAGTATGACTCTGGTGGCCCGTGGGTTTGCAATGGAG  659
         | || |||| |:|| |:|| || :||||||| |:||| ||:||: |||:||| |||||
Qy  302  GNCARNNNCARNNNGAYWSNTGYAAYGNGAYWSNGNGNGNCCNYTNATHTGYAAYGGNT  361
Db  660  AAGTCCAGGTGATTGTTTCCTGGGGTGATG-GC--TGTGCTTTGGAAGGGAAGCCTGGTG  716
         :||||||||| |||||:||:||||| || | ||  |||||:|||||: |||| :|||:|
Qy  362  AYYTNCARGGNYTNGYNWSNTTYGGNAARGCNCCNGYYGGNCARGYNGGNGNTNCCNGNG  421
Db  717  TCTACCACCAAGGTCTGCAACTACCTGAACTGGATTCATCAGAC  759
         |:|| ||| ||| ||:|||| |||:|||| ||||::|| ||:|
Qy  422  TNTAYACNAAYYTNGYAARTTYACNGARTGGATHGARARARAC  464

RESULT  15
LOCUS       MMTAM1            428 bp    mRNA              ROD       10-JUL-1995
DEFINITION  Mouse mRNA for gamma-7S nerve growth factor (y-NGF) fragment.
ACCESSION   X00472
NID         g54260
VERSION     X00472.1    GI:54260
KEYWORDS    complementary DNA; nerve growth factor; serine protease.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 428)
AUTHORS     Howles,P.N., Dickinson,D.P., DiCaprio,L.L., Woodworth-Gutai,M. and
            Gross,K.W.
TITLE       Use of a cDNA recombinant for the gamma-subunit of mouse nerve
            growth factor to localize members of this multigene family near the
            TAM-1 locus on chromosome 7
JOURNAL     Nucleic Acids Res. 12 (6), 2791-2805 (1984)
MEDLINE     84169573
FEATURES    Location/Qualifiers
   source   1..428
            /organism="Mus musculus"
            /db_xref="taxon:10090"
```

```
CDS              <1..408
                 /note="gamma-NGF"
                 /codon_start=1
                 /protein_id="CAA25154.1"
                 /db_xref="PID:g54261"
                 /db_xref="GI:54261"
                 /db_xref="MGD:MGI:97322"
                 /db_xref="SPTREMBL:Q62284"
                 /translation="SKPADITDTVKPITLPTEEPKLGSTCLASGWGSITPTKFQFTDD
                 LYCVNLKLLPNEDCAKAHIEKVTDAMLCAGEMEGGKDTCKGDSGGPLICDGVLQGITS
                 WGHTPCGEPDMPGVTKLNKFTSWIKDTMAKNP"
BASE COUNT      123 a   116 c   102 g    87 t
ORIGIN
```

```
Query Match         19.5%;   Score 365;  DB 32;  Length 428;
Best Local Similarity 45.7%;  Pred. No. 7.62e-22;
Matches   84;  Conservative   28;  Mismatches   72;  Indels    0;  Gaps    0;

Db    206 CCATGCTCGTGTGCAGGAGAGATGGCGAAGGAGGCAAAGACACTTGCAAGGGTGACTCAGGAG 265
          : ||| ||:: ||||  |         |:        || :::| || ||::: ||       |
Qy    281 SNATGTTYTGYCNGGNGGNGGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNG 340

Db    266 GCCCACTGATCTGTGATGGTCTCCAAGGTATCACATCATGGGGCCATACCCCATGCG 325
          | || :|| :||:: |:|  | |  ::| |||| |     | || ||  || ||:|
Qy    341 GNCCNYTNATHTGYAAYGGNTAYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYG 400

Db    326 GTGAACCTGATATGCCAGGCGTCTACCACCAAACTTAATAAGTTACCTCCTGGATAAAAG 385
          | | : | || ||| ||  || ||| | : |  || :|   ||||| :|
Qy    401 GNCARGTNGGNGTNCCNGGNGTNTAYACNAAYTNTGYAARTTYACNGARTGGATHGARA 460

Db    386 ACAC 389
          |  ||
Qy    461 ARAC 464
```

Search completed: Wed Sep 29 01:58:59 1999
Job time : 922 secs.

```
 _   _   ___   ___   ___   ___   _   _
| \ / | | _ \ / __| | _ \ / __| | | | |
|  V  | | _/  \__ \ |   /| (__  | |_| |   (TM)
|_| |_| |_|   |___/ |_|_\ \___|  \___/
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Sep 29 02:14:24 1999; MasPar time 127.97 Seconds
                                    798.701 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172
Description: (1:-159) from US09030606.pep
Perfect Score: 1871
N.A. Sequence: 1 ATGGTNGARGCNWSNYTNWS..........ARAARACNGTNCARGCNWSN 477
                    TACCANCTYCGNWSNRANWS..........TYTTYTGNCANGTYCGNWSN
Comp:

Scoring table: TABLE bktranslate2
                 Gap 30

Nmatch STD :   Dbase 0; Query 0

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Statistics: Mean 47.957; Variance 215.185; scale 0.223
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Database: n-geneseq35
```
1:part1  2:part2  3:part3  4:part4  5:part5  6:part6  7:part7
8:part8  9:part9  10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40 41:part41 42:part42 43:part43
44:part44 45:part45 46:part46 47:part47 48:part48
49:part49 50:part50 51:part51 52:part52 53:part53
54:part54 55:part55 56:part56 57:part57 58:part58
59:part59 60:part60
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	1799	96.2	1248	51	V61249	cDNA sequence of pros	3.96e-150
2	1799	96.0	1248	49	V58644	Prostate tumour speci	3.96e-150
3	1778	95.0	1386	44	V11855	Homo sapiens Tub Inte	3.81e-148
4	1769	94.5	871	44	V37495	Human prostate-specif	2.70e-147
5	1759	94.0	1167	49	V58647	Prostate tumour speci	2.37e-146
6	1759	94.0	1167	51	V61252	cDNA sequence of pros	2.37e-146
7	1105	59.1	1265	49	V58645	Prostate tumour speci	5.32e-85
8	1105	59.1	1265	51	V61250	cDNA sequence of pros	5.32e-85
9	870	46.5	234	51	V61158	cDNA sequence of pros	2.95e-63
10	870	46.5	234	49	V58522	Prostate tumour speci	2.95e-63
11	704	37.6	1459	49	V58646	Prostate tumour speci	4.48e-48
12	704	37.6	1459	51	V61251	cDNA sequence of pros	4.48e-48
13	699	37.4	1119	49	V58648	Prostate tumour speci	1.28e-47
14	699	37.4	1119	51	V61253	cDNA sequence of pros	1.28e-47
15	489	26.1	732	9	Q53487	DNA encoding Zyme APP	8.72e-29
16	489	26.1	1438	32	T79126	Human serine protease	8.72e-29
17	489	26.1	1526	42	V07152	Protease M, a novel s	8.72e-29
18	477	25.5	986	14	Q81203	Human stratum corneum	9.96e-28
19	477	25.5	1089	24	T39783	Human amyloid precurs	9.96e-28
20	332	17.7	833	47	V42925	DNA encoding a human	3.36e-15
21	332	17.7	1146	56	V84589	Human secreted protei	3.36e-15
22	323	17.3	90	51	V61208	cDNA sequence of pros	1.92e-14
23	323	17.3	90	49	V58593	Prostate tumour speci	1.92e-14
24	316	16.9	699	39	V16367	Nucleotide sequence o	7.43e-14
25	311	16.6	693	50	V61861	A. contortrix protein	1.94e-13
26	306	16.4	1003	21	T13316	Korean viper salmosa	5.08e-13
27	290	15.5	741	1	N81633	Human spleen trypsin	1.08e-11
28	290	15.5	744	17	T04000	Human pancreatic tryp	1.08e-11
29	290	15.5	744	17	T04001	Human pancreatic tryp	1.08e-11
30	290	15.5	744	17	T03999	Human pancreatic tryp	1.08e-11
31	286	15.3	957	3	Q20501	Encodes fibrinogenoly	2.30e-11
32	284	15.2	1333	27	T48519	Human neuropsin-encod	3.36e-11
33	284	15.2	1333	29	T63251	Mouse neuropsin gene.	3.36e-11
34	280	15.0	1078	38	T91308	Human H83-22 secreted	7.16e-11
35	275	14.7	1523	1	N81741	Batroxobin gene	1.84e-10
36	275	14.7	1523	1	Q04757	Sequence encoding bat	1.84e-10
37	274	14.6	790	45	V24548	Trypsinogen-like prot	2.22e-10
38	274	14.6	944	59	V84052	Nucleic acid encoding	2.22e-10
39	266	14.2	988	35	Q20500	Encodes fibrinogenoly	9.96e-10
40	263	14.1	992	35	T91054	Human prostate specif	1.75e-09
41	263	14.1	1445	44	V32497	Prostate specific ant	1.75e-09
42	263	14.1	1462	35	T91055	Human prostate specif	1.75e-09
43	263	14.1	1466	44	V32496	Prostate-specific ant	1.75e-09
44	263	14.1	1728	22	T35867	Prostate-specific ant	1.75e-09
45	263	14.1	1729	15	T04864	Prostate-specific ant	1.75e-09

ALIGNMENTS

```
RESULT  1
ID   V61249 standard; cDNA; 1248 BP.
AC   V61249;
DT   06-JAN-1999  (first entry)
DE   cDNA sequence of prostate tumour clone P703 splice variant DE1
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03492.
PR   09-FEB-1998;  US-020956.
PR   25-FEB-1997;  US-806099.
PR   01-AUG-1997;  US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609986/51.
DR   P-PSDB: W71871.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 104; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 1248 BP;  288 A;  424 C;  303 G;  228 T;

Query Match             96.2%;  Score 1799;  DB 51;  Length 1248;
Best Local Similarity   55.5%;  Pred. No. 3.96e-150;
Matches  264;  Conservative 115;  Mismatches 97;  Indels 0;  Gaps 0;
```

```
Db   217 atggtggaggcagcctcctccgtacggcaccagagtacaacagacccttgctcgctaac 276
         |||||||||| ::|| ||  ::|||||||||  :|||::|:||  || ::|| || ||||
Qy     1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY 60

Db   277 gacctcatgctctcaagttgacgaatccgtctcgagttcgacaccatccggagcatc 336
         |||||| ||||| ||:||| ||  :||||||||   ::|||  |||::| :: :|
Qy    61 GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   337 agcattgcttcgcagtgcccaccgcgggaactcttgcctcgtttctgctggggtctg 396
         :: :|::|| ||||||  || |||| |||||||| ||||:|| |||||:| ::|
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db   397 ctggcgaacggcagaatgctaccgtgctgcagtgcgtgcagtgtcggtggtgtctgag 456
         ::|| ||||| :|  || ||:| |||| ||:|||||||||||||||||| |||||||||
Qy   181 YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db   457 gaggtctgcagtaagctctatgacccgctgtaccacccccagcatgttctgcgccgcgga 516
         |||||||||||||||| ||||  ||||||||||||  |||||||| |||  |:|||| ||
Qy   241 GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYCNGCNGGN 300

Db   517 gggcaagaccagaaggactcctgcaacggtgactctggggggccctgatctgcaacggg 576
         |||| ::  |  :|||||| ||||||  ::||||||| ::||  ||||| |||| |||||
Qy   301 GGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGN 360

Db   577 tacttgcaggcgccttgtcttgcccggaaaagcccgtgtggccaagttggcgtgccaggt 636
         |||||||||||  ||||| :||:|  :||||||||  :||:| ||||||| |:|||||||
Qy   361 TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 420

Db   637 gtctacccaacctctgcaaattcactgagtgagtagagaaaccgtccaggccag 692
         |||||||| ||| ||||||| ||||||| | ||| |||||  |||||||||||| ||
Qy   421 GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476
```

```
Db   217 atggtggaggcagcctcctccgtacggcaccagagtacaacagacccttgctcgctaac 276
         |||||||||| ::|| ||  ::|||||||||  :|||::|:||  || ::|| || ||||
Qy     1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY 60

Db   277 gacctcatgctctcaagttgacgaatccgtctcgagttcgacaccatccggagcatc 336
         |||||| ||||| ||:||| ||  :||||||||  :: |||  |||::| :: :|
Qy    61 GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   337 agcattgcttcgcagtgcccaccgcgggaactcttgcctcgtttctgctggggtctg 396
         :: :|::|| ||||||  || |||| |||||||| ||||:|| |||||:| :||||| :|
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db   397 ctggcgaacggcagaatgctaccgtgctgcagtgcgtgcagtgtcggtggtgtctgag 456
         ::|| ||||| :|  || ||:| |||| ||:|||||||||||||||||| |||||||||
Qy   181 YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db   457 gaggtctgcagtaagctctatgacccgctgtaccacccccagcatgttctgcgccgcgga 516
         ||||||||||||||| |||| :||||||||||||  |||||||| || :|:|||| ||
Qy   241 GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYCNGCNGGN 300

Db   517 gggcaagaccagaaggactcctgcaacggtgactctggggggccctgatctgcaacggg 576
         |||| ::  |  :|||||| ||||||  ::||||||| ::||  ||||| |||| |||||
Qy   301 GGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGN 360

Db   577 tacttgcaggcgccttgtcttgcccggaaaagcccgtgtggccaagttggcgtgccaggt 636
         |||||||||||  ||||| :||:|  :||||||||  :||:| ||||||| |:|||||||
Qy   361 TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 420

Db   637 gtctacccaacctctgcaaattcactgagtgagtagagaaaccgtccaggccag 692
         |||||||| ||| ||||||| ||||||| | ||| |||||  |||||||||||| ::
Qy   421 GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476
```

```
Db   217 atggtggaggcagcctcctccgtacggcaccagagtacaacagacccttgctcgctaac 276
```

CC drug screening.
SQ Sequence 1386 BP;   318 A;   490 C;   321 G;   249 T;

Query Match   95.0%;   Score 1778;   DB 44;   Length 1386;
Best Local Similarity   55.5%;   Pred. No. 3.81e-148;
Matches 262; Conservative 113; Mismatches 97; Indels 0; Gaps 0;

Db   197 atggtgaggccagcctccgtacggcaccagagtacaacagacccttgtctgctaac 256
          |||||| ||||| : ||| :|:||::|| ||:||::|::|::|||::| :| :||:|
Qy   1 ATGTGNARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY 60

Db   257 gacctcatgtctcatcaagttggacgaatccgtccgagtctgacaccatccggagcatc 316
          ||:| ||:||| |:||||| :||||:::|||:||||::| ||:||:::|::||||:||
Qy   61 GAYYTNATGTVNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   317 agcattgcttcgcagtgcctaccgcgggagaactcttgcctgtttctgctgggtctg 376
          :: ||||||:: :|||||:||||| |||:| | :||||:||:| ::||| ::||| :|
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTN 180

Db   377 ctggcgaacggcagaatgcctaccgtgtcgagtgcgtgaacgtgtcggtggtctgag 436
          :| ||::|||:: :|||||:||||| ||:||:||:|| ||:||:||::|| ||:||:||
Qy   181 YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db   437 gaggtctcagtaagctctatgacccgctgtaccaccccagcagtcttcgcgcgcgga 496
          ||||||:||:||:: |||:||||| || ||:||:|| :::|||:||:||:||:||:|||
Qy   241 GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGN 300

Db   497 gggcaagaccagaggactcctgcaacgtgactctgggggcccctgacctgcaacggg 556
          |||:||||||||:||||:||||| ||:|||:||||||||:|||:||:||:||:||:|||
Qy   301 GGNCARNNNCARNNNGAYWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 420

Db   557 tacttgcaggggcctgtgtctttcggaaaagccgtgtggccaagtggccgtgccaggt 616
          ||:||:|||||:||||:||||:|:|||:|||:||||:||||:||||||:||:|||||:||
Qy   361 TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 472

Db   617 gtctacaccaacctctgcaaattcactgagtggatagagaaaccgtaccag 668
          |:|||:||:|::||:|:|:|| ||:|||:||||:||:|||:|:| :|::|
Qy   421 GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARARARACNGTNCARG 472

RESULT 4
ID   V37495 standard; DNA; 871 BP.
AC   V37495;
DT   07-SEP-1998 (first entry)
DE   Human prostate-specific kallikrein (HPSK) encoding DNA.
KW   Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW   benign prostate hyperplasia; diagnosis; drug screening; PSK; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   31..777
FT   /tag= a
FT   /transl_except= (pos:367..369, aa:Xaa)
FT   /transl_except= (pos:412..414, aa:Xaa)
FT   /transl_except= (pos:424..426, aa:Xaa)
FT   /product= "HPSK protein"
FT   /note= "Xaa = unknown"
PN   WO9820117-A1.
PD   14-MAY-1998.
PF   31-OCT-1997; U20051.
PR   05-NOV-1996; US-744026.
PA   (INCY-) INCYTE PHARM INC.
PI   Bandman O, Goli SK;
DR   WPI; 98-286933/25.
DR   P-PSDB; W60592.
PT   New isolated prostate-specific kallikrein - used to develop products
PT   for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT   hyperplasia
PS   Claim 5; Fig 1A-C; 68pp; English.
CC   This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC   containing an expression vector comprising the HPSK nucleic acid sequence
CC   can be used to produce the protein recombinantly. The HPSK products can
CC   be used for the diagnosis of conditions or diseases associated with
CC   expression of HPSK such as prostate carcinoma and benign prostate

CC   hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC   modulate its activity can be used for the preparation of treatment of
CC   such conditions or diseases. The products can also be used for detection
CC   and drug screening, especially for the detection of prostate-specific
CC   kallikrein (PSK).
SQ   Sequence 871 BP;   166 A;   260 C;   258 G;   184 T;

Query Match   94.5%;   Score 1769;   DB 44;   Length 871;
Best Local Similarity   55.3%;   Pred. No. 2.70e-147;
Matches 263; Conservative 113; Mismatches 100; Indels 0; Gaps 0;

Db   298 atgtggaggccagcctctccgtacggcaccagagtacaacagacccttgtctcgctaac 357
          ||||| |||||| ||| : ||| :|:||::|| ||:||::|::|::| ||::| :||:|
Qy   1 ATGTGNARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY 60

Db   358 gacctcagtcatcaagttggacgaatcctgtcgcgagtctgacaacatccggagnatc 417
          ||:| ||:| |:||||| :||||:::|| :|:::||||::||:||:::|::||| :||
Qy   61 GAYYTNATGTYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   418 agcattgttcgcagtgcctaccgcggggaactttgcctgtttctggctgggtctg 477
          :: ||||::|:: :|||||:||||| |||:|| :::||||:||:| ::||| ::||| :|
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTN 180

Db   478 ctggcgaacggcagaatgcctaccgtctgcagtcgtcgcagtggtctgag 537
          :| ||::|||:: :|||||:||||| ||:|| |::||:||::|| ||::|||
Qy   181 YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db   538 gaggtctcagtaagctctatgacccgctgtaccaccccagcagtctctgcgccgcgga 597
          ||||||:||:||:: |||:||||| || ||:||:|| :::|||:||::| ||:||:|||
Qy   241 GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGN 300

Db   598 gggcaagaccagaggactcctgcaacgtgactcctggggggcccctgatctgcaacggg 657
          |||:||||||||:||||:||||| ||:|||:||||||||:|||:||:||:||:||:|||
Qy   301 GGNCARNNNCARNNNGAYWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 360

Db   658 tacttgcaggggcctgtgtcttttcggaaaagccgtgtgtggccaagttggcgtgccaggt 717
          ||:||:|||||:||||:||||:|:|||:|||:||||:| :|||:||||||:||:|||||:||
Qy   361 TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 420

Db   718 gtctacaccaacctctgcaaattcactgatgatggatagagaaaccgtccaggccag 773
          |:|||:||:|::||:|:|:|| ||:|||::||:||:|||:|||:|:|| :|::::|
Qy   421 GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARARARACNTNCARGCNWS 476

RESULT 5
ID   V58647 standard; cDNA; 1167 BP.
AC   V58647;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone DE13.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   28..645
FT   /*tag= a
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   P-PSDB; W69388.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 115; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic

Right column

```
Matches  262; Conservative  113; Mismatches 101; Indels   0; Gaps   0;

Db  166 atggtggaggccagcctccgtacggcaccagagtacaacagactcttgctcgctaac 225
         |||||:|||| :|: :|  ::| |:|||:|| ||:::|| ||:: :|:| | |:||:: |
Qy    1 ATGTGNGARGCNWSNYTNWSNGTNWSNGNCAYCCNGARTAYAAYMGNCCNYTNYNGCNAAY 60

Db  226 gacctcatgtctcatcaagttggacgaatccgtgtccagtctgacacccatccggagcatc 285
         |:|:|||  :||:||: ||||  | | ||  ||:|| |||| :: :|| ||| :|| | |:
Qy   61 GAYYTNATGTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db  286 agcattgcttcgcagtgcctaccgcggggaactcttcctcgtntctggctgggtggtcg 345
         :: |||:|| :: ||||  :|| ||| ||  | |||  |:|:|| |||||:  || | ::|
Qy  121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTN 180

Db  346 ctggcgaacggcagaatgcctaccgtgctgcactgcgtgaacgtgtcggtgtgtctgag 405
         |:||| ||  | |:  | :|| || :|:| ||||:|| |: :|:||:|| || |:||:||
Qy  181 YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR 240

Db  406 gangtctgcagtaagctctatgaccgcgtaccacccagcatgttctgcgcggcgga 465
         || |||||||| |:|:|| :|| ||:|:||  || :||: || || ||:||:||:|||
Qy  241 GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCACCCNWSNATGTTYTGYCNGGNGGN 300

Db  466 gggcaagaccagaaggacctcctgcaacgtgctgactctggggcccctgatctgcaacggg 525
         ||| || | || |:| :|| :||||| :| :||||| || ||| ||:|| | |:|| || |
Qy  301 GGNCARNNNCARNNNGAYWSNTGYAAYGTNWSNGGNGGNCCNYTNATHTGYAAYGGN 360

Db  526 tacttgcagggcttgtgtcttcggaaaagcccgtgtggccaacttggcgtgccagt 585
         || :||||| |||:|:| :| |||||||| ||||| || ||| ||| | :|| || |
Qy  361 TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGN 420

Db  586 gtctacaccaacctctgcaaattcactgagtgatagaaaaccgtccagnccag 641
         |||:||:|| || ||||| ||:| || || ||: | |||| |:|||:|||
Qy  421 GTNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476
```

```
RESULT 7

ID   V58645 standard; cDNA; 1265 BP.
AC   V58645;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone DE2.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO983741B-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 113-114; 14lpp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence  1265 BP;  256 A;  432 C;  321 G;  245 T;

Query Match     59.1%;  Score 1105;  DB 49;  Length 1265;
Best Local Similarity  58.2%;   Pred. No. 5.32e-85;
Matches  167; Conservative  62; Mismatches  58; Indels   0; Gaps   0;

Db  450 ggcagaatgcctaccgtgctgcgagtcgtcgagtgtgtctgaggaggtctgc 509
         || :| |||| ||||| :|:| ||:|| :| ||:| ||:|| || ||:||| |
Qy  190 GGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGY 249
```

Left column

```
RESULT 6

ID   V61252 standard; cDNA; 1167 BP.
AC   V61252;
DT   06-JAN-1999 (first entry)
DE   cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO983709-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
DR   P-PSDB; W71782.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 107; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence  1167 BP;  242 A;  400 C;  287 G;  222 T;

Query Match     94.0%;  Score 1759;  DB 51;  Length 1167;
Best Local Similarity  55.0%;   Pred. No. 2.37e-146;
```

```
Db   510 agtaagctctatgaccgtgtaccacccagcatgttctgcgccggcggaggcaagac 569
          ::|:||:||:| ||:||:|  ||:||::  ||||:||  ||:||:||
Qy   250 WSNAARYTNTAYGAYCCNWSNATGTTYTGYGCNGGNGNGGNCARNNN 309

Db   570 cagaaggactcctgcaacggtacttgcag 629
          ||  ::  |||:||||::||  ||||  ||
Qy   310 CARNNNGAYWSNTGYAAYGGNGAYWSNTGYAAYGGNGAYTNTGYAAYGGNTAYYTNCAR 369

Db   630 ggccttgtgctttcggaaaagccccgtggccagtggcgttcacctgttctcacac 689
          ||:||:||  ::  ||  :||:|||||
Qy   370 GGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNTNCCNGNGTNTAYACN 429

Db   690 aacctctgcaaattcactgagtggatagagaaaaccgtccaggccag 736
          ||:|| || ||:|| || ||:|||:||  :||:|||||
Qy   430 AAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476

RESULT 8
ID   V61250 standard; cDNA; 1265 BP.
AC   V61250;
DT   06-JAN-1999 (first entry)
DE   cDNA sequence of prostate tumour clone P703 splice variant DE2.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1998; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 105-106; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 1265 BP;  256 A;  432 C;  321 G;  245 T;

Query Match          59.1%;  Score 1105;  DB 51;  Length 1265;
Best Local Similarity 58.2%;  Pred. No. 5.32e-85;
Matches  167;  Conservative  62;  Mismatches  58;  Indels  0;  Gaps  0;

Db   450 ggcagaatgctaccgtgctgcagtgcggtgaacgtgtcgtggtgtctgaggaggtctgc 509
          ||  :: ||| || ||| |||:|||   ||:|||:|| ::  |||:|:|| ||:|||
Qy   190 GGNMGNATGCCNACNGTNYTNCARTGYGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGY 249

Db   510 agtaagctctatgaccgtgtaccacccagcatgttctgcgccggcggaggcaagac 569
          ::|:||:||:| ||:||:|  ||:||::  ||||:||  ||:||:||
Qy   250 WSNAARYTNTAYGAYCCNWSNATGTTYTGYGCNGGNGNGGNCARNNN 309

Db   570 cagaaggactcctgcaacggtacttgcag 629
          ||  :: |||:|||||::||  ||||  ||
Qy   310 CARNNNGAYWSNTGYAAYGGNGAYWSNTGYAAYGGNGAYTNTGYAAYGGNTAYYTNCAR 369

Db   630 ggccttgtgctttcggaaaagccccgtggccagtggcgttcacctgttctcacac 689
          ||:||:||  ::  ||  :||:|||||
Qy   370 GGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNTNCCNGNGTNTAYACN 429

Db   690 aacctctgcaaattcactgagtggatagagaaaaccgtccaggccag 736
          ||:|| || ||:|| || ||:|||:||  :||:|||||
Qy   430 AAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476

RESULT 9
ID   V61168 standard; cDNA; 234 BP.
AC   V61168;
```

```
DT   06-JAN-1999 (first entry)
DE   cDNA sequence of prostate tumour clone P20.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1998; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 53-54; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence  234 BP;  43 A;  68 C;  68 G;  55 T;

Query Match          46.5%;  Score 870;  DB 51;  Length 234;
Best Local Similarity 52.6%;  Pred. No. 2.95e-63;
Matches  123;  Conservative  63;  Mismatches  48;  Indels  0;  Gaps  0;

Db     1 acaacagaccctgctcgctaacgacctcatgtctcatcaagttgacgaatccgtgtccg 60
           |::||:||:|| || ||::|| || || |||:||| |:|| ::||: || :|:||
Qy    38 AYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNG 97

Db    61 agtctgacacatccggagcatcagcattgcttcgcagtgccctaccgcgggaactctt 120
           :: |||:|| || ::||| || ||  ||| :: || ::|| ::|| :|| ||:||:
Qy    98 ARWSNGAYACNATHMGNWSNATHWSNATHGCNWSNCARTGYCCNACNCGNGGNAAYWSNT 157

Db   121 gcctcgtttctgctggtggttgtgctgcgaacggcagaatgctaccgtgctgcagtgcg 180
           :: || :|||:|| ||:|:|| ||| || ::|| ::|| ||| || ||| |||:|||
Qy   158 GYYTNGTNWSNGGNTGGGNYTNYTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYG 217

Db   181 tgaacgtgtcgtggtgtctgaggaggtctgcagtaagtctcatgacccgctgt 234
           |:|||:|| :: |||:|:|| ||:||| ||:|| ::||| || ::|||:||:|
Qy   218 TNAAYGTNWSNGGNTGNTGNWSNGARGARGTNTGYWSNAARYTNTAYGAYCCNYTNT 271

RESULT 10
ID   V58522 standard; cDNA; 234 BP.
AC   V58522;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone P20.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03590.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 56; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
```

Qy   181 YTNGCNAAYGG 191

RESULT   12
ID   V61251 standard; cDNA; 1459 BP.
AC   V61251;
DT   06-JAN-1999 (first entry)
DE   cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO983709-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 106; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 1459 BP;   427 A;   328 C;   406 G;   295 T;

Query Match   37.6%; Score 704; DB 51; Length 1459;
Best Local Similarity   51.8%; Pred. No. 4.48e-48;
Matches   99; Conservative   53; Mismatches   39; Indels   0; Gaps   0;

Db   99 atggtgaggccagcctctccgtacggcaccagagtacaacagacccttgctcgctaac 158
             | |||| ||:|| .|| .| :|:|| | || |||:::::|||::||::|:|||||
Qy   1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY   60

Db   159 gacctcatgctcatcaagttgacgaatccgtgtccagtcgtgacaccatccggagcatc 218
             | |||| : :||||||| ||| || || | :| :||||||| : ||||:| |||::
Qy   61 GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGTNWSNGAYACNATHMGNWSNATH 120

Db   219 agcattgcttcgcagtgccctaccgcgggaaactcttggctgttctggctggggtctg 278
             || |||:| :| |||:| | :|:|||| :|| ||||||:||||||:|:|||:|||:|
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db   279 ctggcgaacgg 289
             :| || ||:||
Qy   181 YTNGCNAAYGG 191

RESULT   13
ID   V58648 standard; cDNA; 1119 BP.
AC   V58648;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone DE14.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   34..528
FT   /*tag= a
PN   WO983709-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   P-PSDB; W69389.
PT   Novel human prostate specific tumour protein and fragments - useful

───

CC   cancers.
SQ   Sequence 234 BP;   43 A;   68 C;   68 G;   55 T;

Query Match   46.5%; Score 870; DB 49; Length 234;
Best Local Similarity   52.6%; Pred. No. 2.95e-63;
Matches   123; Conservative   63; Mismatches   48; Indels   0; Gaps   0;

Db   1 acaacagacccttgctcgctaacgacctcatgctcatcaagttgacgaatccgtgtccg   60
             |||:::| || | || || :|| : |||||||||||| ||:|| || | :| ::|:|
Qy   38 AYAAYMGNCCNYTNYTNGCNAAYGAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNG   97

Db   61 agtctgacaccatccggagcatcagcattgcttcgcagtgccctaccgcgggaaactctt 120
             || :||||| : |||:| |||:| || |||:| :| |||:| | :|:|||| :|| ||
Qy   98 ARWSNGAYACNATHMGNWSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNT 157

Db   121 gcctcgtttctgctggcgggtctgctggcggaacggcagaatgctaccgtgctcagtcg 180
             |:||| | :|||| || || |||:|||||:|| :|| || |||:| |||:| |||:|
Qy   158 GYYTNGTNWSNGGNTGGGGNYTNYTNGCNAAYGGNATGGCNACNGTNYTNCARTGYG 217

Db   181 tgaacgttcggtggtgtctgagaggtctgcagtaagctcatgaccgctgt 234
             | |||:|| :|| :| :|:|| :| | |:||||::|| |:|| |:|:| :|
Qy   218 TNAAYGTNWSNGGNGTNWSNGARGARGYNTGYWSNAARYTNTAYGAYCCNYTNT 271

RESULT   11
ID   V58646 standard; cDNA; 1459 BP.
AC   V58646;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone DE6.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO983709-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 114; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 1459 BP;   427 A;   328 C;   406 G;   295 T;

Query Match   37.6%; Score 704; DB 49; Length 1459;
Best Local Similarity   51.8%; Pred. No. 4.48e-48;
Matches   99; Conservative   53; Mismatches   39; Indels   0; Gaps   0;

Db   99 atggtgaggccagcctctccgtacggcaccagagtacaacagacccttgctcgctaac 158
             | |||| ||:|| .|| .| :|:|| | || |||:::::|||::||::|:|||||
Qy   1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY   60

Db   159 gacctcatgctcatcaagttgacgaatccgtgtccagtcgtgacaccatccggagcatc 218
             | |||| : :||||||| ||| || || | :| :||||||| : ||||:| |||::
Qy   61 GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGTNWSNGAYACNATHMGNWSNATH 120

Db   219 agcattgcttcgcagtgccctaccgcgggaaactcttggctgttctggctggggtctg 278
             || |||:| :| |||:| | :|:|||| :|| ||||||:||||||:|:|||:|||:|
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGNYTN 180

Db   279 ctggcgaacgg 289
             :| || ||:||

PT   for detecting and treating prostate cancers
PS   Claim 1; Page 116-117; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match   37.4%; Score 699; DB 49; Length 1119;
Best Local Similarity   51.6%; Pred. No. 1.28e-47;
Matches   98; Conservative   53; Mismatches   39; Indels   0; Gaps   0;

Db   172 atgggtgaggccagccctccgtacggcaccagagtacaacagaccctgctcgctaac 231
                !!!!!!! !!! ::!:: ! ::!! !!!! !!!!!!!!:! ::!!!!!:!! ::!!
Qy   1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY 60

Db   232 gacctcatgctcatcaagttggacgaatccgtgtccgagtccgaccatccggagcatc 291
                !:!!:! !!!! :! !!! ::! !!!!!:! ::!! :! !!!!!!!!!:! !!:
Qy   61 GAYYTNATGYTNATCAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   292 agcattgcttcgcagtgccctaccgcgggagaacttgctcttcgtttctggtggggtctg 351
                ::!!:!!:! !!!! !!! !!!! !!!!!! !!:! !!!!!! !!!:! !!!!! :!
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNGYYTNGTNWSNGGNTGGGGNYTN 180

Db   352 ctggcgaacg 361
                !!:! :!!:!
Qy   181 YTNGCNAAYG 190

RESULT   14
ID   V61253 standard; cDNA; 1119 BP.
AC   V61253;
DT   06-JAN-1999 (first entry)
DE   cDNA sequence of prostate tumour clone P703 splice variant DE14.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
WPI; 98-609886/51.
DR   P-PSDB; W71873.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 108-109; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match   37.4%; Score 699; DB 51; Length 1119;
Best Local Similarity   51.6%; Pred. No. 1.28e-47;
Matches   98; Conservative   53; Mismatches   39; Indels   0; Gaps   0;

Db   172 atgggtgaggccagccctccgtacggcaccagagtacaacagaccctgctcgctaac 231
                !!!!!!! !!! ::!:: ! ::!! !!!! !!!!!!!!:! ::!!!!!:!! ::!!
Qy   1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY 60

Db   232 gacctcatgctcatcaagttggacgaatccgtgtccgagtcgaccatccggagcatc 291
                !:!!:! !!!! :! !!! ::! !!!!!:! ::!! :!
Qy   61 GAYYTNATGYTNATCAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   292 agcattgcttcgcagtgccctaccgcgggagaacttgctcttcgtttctggtggggtctg 351
                ::!!:!!:! !!!! !!! !!!! !!!!!! !!:! !!!!!! !!!:! !!!!! :!
Qy   121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNGYYTNGTNWSNGGNTGGGGNYTN 180

Db   352 ctggcgaacg 361
                !!:! :!!:!
Qy   181 YTNGCNAAYG 190

RESULT   15
ID   Q53487 standard; DNA; 732 BP.
AC   Q53487;
DT   01-JUL-1994 (first entry)
DE   DNA encoding Zyme APP-cleaving protease.
KW   Amyloid precursor protein-cleaving protease; Alzheimer's disease;
KW   Down's syndrome; diagnosis; propensity; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   cds   1..732
FT   /*tag= a
PN   EP-576152-A.
PD   29-DEC-1993.
PF   26-MAY-1993; 304103.
PR   28-MAY-1992; US-891542.
PA   (ELIL) LILLY & CO ELI.
PI   Dixon EP, Johnstone EM, Little SP, Norris FH;
WPI; 94-001306/01.
DR   P-PSDB; R44532.
PT   Amyloid precursor protein-cleaving protease - associated with
PT   Alzheimer's disease and Down's syndrome
PS   Claim 3; Page 13; 14pp; English.
CC   The sequence is that encoding Zyme, an amyloid precursor protein-
CC   cleaving protease. It can be used to diagnose Alzheimer's disease
CC   or a propensity to develop Alzheimer's disease, by digesting the
CC   patient's DNA, hybridising with it and comparing the pattern to
CC   those members of the patient's family who exhibit the disease. It
CC   can also be used to diagnose Down's syndrome and is useful for
CC   determining whether a test substance is a functional ligand for
CC   the Zyme APP-cleaving protease.
SQ   Sequence 732 BP; 168 A; 212 C; 203 G; 149 T;

Query Match   26.1%; Score 489; DB 9; Length 732;
Best Local Similarity   36.2%; Pred. No. 8.72e-29;
Matches 165; Conservative   71; Mismatches 220; Indels   0; Gaps   0;

Db   275 ctgtgatccaccctgactatgatgccgcagccatgaccaggacatcatgctgttcgcc 334
                ! !! :! !!!! !!!:! !! !::! !!! !!:! !!!:! !!!!!! !:!
Qy   20 SNGTNMGNCAYCCNGARTAYAAYAAYMGNCCNYTNYTNGCNAAYGAYTNATGYTNATHAARY 79

Db   335 tggcacgcccagccaaactctgaactcatccagccctncccctgagagggactgct 394
                !! :!!!! !!:! !! :::! !!! !!!:! !!! :::!! !! !!!:
Qy   80 TNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNAICHGCNWSNCARTGYC 139

Db   395 cagccaacaccagctgccacatcctggctgggcaagacagcagatggtgatttcc 454
                !! :::! !!!! !! :! !!! !!!:! !!! :::!! !!! !!!:
Qy   140 CNACNGCNGGNAAYWSNGYYTNGTNWSNGGNTGGGGNYTNYTNGCNAAYGGNMGNATGC 199

Db   455 ctgaccatccagtgtgcatacatccacctgtgtccgtgagcatcacctgttgagcatgcct 514
                ! !!!! !! :! !!!!:! !! :!!:! !!! !!! :::! !! !!!! !:!
Qy   200 CNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNT 259

Db   515 acctggccagatcaccccagaacatgttgtgtgctgggatgaagtacgggaaggatt 574
                !!!!! !! :::! !!!! :! !! :::!! !!! :! !!! :::! !!!:
Qy   260 AYGAYCCNYTNTAYCAYCCNWSNATGTTYGTGYGCNGNGGNGNGNGCARNNNCARNNNGAYW 319

Db   575 cctgccagggtgattctggggtccgtcggtatgtggagaccacctccgaggccttgt 634
                !! :::!! !!!!:! !!! :! !!! :! !!!! !!! :::! !!!:
Qy   320 SNTGYAAYGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYTNCARGGNYTNGTNW 379

Db   635 catgggtaacatccccctggatcaaaggagagagcaggaggatcatcgca 694
                !! !! :::!! !!!!:! !!! :! !!! :! !!!!
Qy   380 SNTTYGGNAARGCNCCNCCNTGYGYGNCARGTNGGNTNCCNGGNGTNTAYACNAAYYTNTGYA 439

US-09-030-606-172.rng

Wed Sep 29 14:24:00 1999

Db   695 gatacacgaactggatccaaaaaaccattcaggcca 730
          .:.::|| | |||||: |:||:|| | |!:|| .:
Qy   440 ARTTYACNGARTGDATHGARAARACNGTNCARGCNW 475

Search completed: Wed Sep 29 02:16:38 1999
Job time : 134 secs.

```
*********************************************************
*********************************************************
*********************************************************
     MPSRCH (TM)
*********************************************************
*********************************************************
*********************************************************
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_tpn n.a. - n.a. Smith-Waterman search, using a protein query
which has been backtranslated into n.a. using IUPAC symbols

Run on: Wed Sep 29 01:59:18 1999; MasPar time 874.18 Seconds
1278.562 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172
Description: (1-159) from US09030606.pep
Perfect Score: 1871
N.A. Sequence: 1 ATGGTNGARGCNWSNYTNWS..........ARAARACNGTNCARGCNWSN 477
Comp: TACCANCTYCGNWSNRANWS..........TYTYTGNCANGTYCGNWSN

Scoring table: TABLE bktranslate2
Gap 30

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 seqs, 1171580779 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: embl-est58
1:em_est10 2:em_est11 3:em_est17 4:em_est18 5:em_est2
6:em_est9 7:em_gss1
Database: genbank-est111
8:gb_est1 9:gb_est10 10:gb_est11 11:gb_est12 12:gb_est13
13:gb_est14 14:gb_est15 15:gb_est16 16:gb_est17
17:gb_est18 18:gb_est19 19:gb_est20 20:gb_est20
21:gb_est21 22:gb_est22 23:gb_est23 24:gb_est24
25:gb_est25 26:gb_est26 27:gb_est27 28:gb_est28
29:gb_est29 30:gb_est3 31:gb_est4 32:gb_est5 33:gb_est6
34:gb_est7 35:gb_est8 36:gb_est9 37:gb_gss1 38:gb_gss2
39:gb_gss3 40:gb_gss4 41:gb_gss5 42:gb_gss6

Statistics: Mean 62.729; Variance 68.475; scale 0.916

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
	1	1452	77.6	777	28	AI557281	PT2.1_15_G12.r tumor2	0.00e+00
	2	1392	74.4	415	14	AA551449	nj55e05.s1 NCI_CGAP_Pr	0.00e+00
	3	611	32.7	576	14	AA533140	nj46h06.s1 NCI_CGAP_Pr	1.65e-157
	4	608	32.5	722	28	AI557025	PT2.1_10_F05.r tumor2	1.95e-156
c	5	474	25.3	404	14	AA503963	nh39a01.s1 NCI_CGAP_Pr	3.25e-109
	6	443	23.7	457	28	AI385433	m185b05.y1 Soares mous	1.55e-98
	7	439	23.5	498	34	W73168	zd55e11.r1 Soares_feta	3.63e-97
	8	439	23.5	505	26	AI391329	mb71h07.y1 Soares mous	3.63e-97
	9	428	22.9	517	10	AA293027	zt54a12.r1 Soares ovar	2.07e-93
c	10	419	22.4	686	26	AI415008	mb71h07.x1 Soares mous	2.38e-90
	11	407	21.8	590	13	AA411252	zt33b03.r1 Soares ovar	2.76e-86
c	12	406	21.7	586	34	W73140	zd55e11.s1 Soares_feta	6.01e-86
	13	381	20.4	625	13	AA419284	zv35b02.r1 Soares ovar	1.48e-77
c	14	380	20.3	496	23	AA864127	vn42g07.r1 Stratagene	3.20e-77
	15	368	19.7	456	23	AI139437	qc20e03.x1 Soares_feta	3.13e-73
	16	367	19.6	392	33	W16362	mb56d06.r1 Soares mous	6.71e-73
	17	356	19.0	400	10	AA293231	zt26g09.r1 Soares ovar	2.88e-69
c	18	356	19.0	539	13	AA477689	zu44a12.r1 Soares ovar	2.88e-69
c	19	347	18.5	551	36	AA101044	zm27el1.s1 Stratagene	2.60e-66
c	20	338	18.1	181	20	AA918913	ol65c02.s1 NCI_CGAP_K1	2.26e-63
c	21	337	18.0	460	18	AA791893	vs54h07.r1 Stratagene	4.77e-63
c	22	335	17.9	493	25	AI324874	m185b05.x1 Soares mous	2.13e-62
c	23	330	17.6	247	36	AA087110	mk20h02.r1 Soares mous	8.93e-61
c	24	330	17.6	585	14	C23111	C23111 Japanese flound	8.93e-61
c	25	327	17.5	386	26	AI110984	EST239277 Normalized r	8.35e-60
c	26	327	17.5	456	22	AI044130	UI-R-C1-jv-h-04-0-UI.s	8.35e-60
	27	326	17.4	1029	33	W13212	ma83a07.r1 Soares mous	1.76e-59
	28	322	17.2	470	23	AU052038	AU052038 Cyprinus carp	3.42e-58
	29	317	16.9	453	23	AI169562	EST215440 Normalized r	1.39e-56
c	30	316	16.9	522	20	AA880451	vw89h02.r1 Stratagene	2.90e-56
c	31	315	16.8	193	35	AA062066	mj183d11.r1 Soares mous	6.06e-56
	32	308	16.5	420	24	AI226226	ue88f06.y1 Soares mous	1.04e-53
c	33	306	16.4	163	12	AA394040	zt52h01.s1 Soares ovar	4.53e-53
	34	306	16.4	467	30	AA073833	mj99h09.r1 Soares mous	4.53e-53
c	35	305	16.3	483	23	AI183346	qd41b10.x1 Soares_feta	9.41e-53
c	36	302	16.1	367	21	AI002101	ot38d03.s1 Soares_test	8.43e-52
c	37	301	16.1	470	15	AA562966	v157d05.r1 Stratagene	1.75e-51
c	38	302	16.1	480	20	AA846771	aj41f01.s1 Soares_test	8.43e-52
c	39	302	16.1	558	17	AA726122	vu88c02.r1 Stratagene	8.43e-52
c	40	300	16.0	498	26	AA925291	UI-R-A1-ee-h-08-0-UI.s	3.62e-51
c	41	297	15.9	324	23	AI177474	EST221106 Normalized r	3.22e-50
c	42	297	15.9	586	17	AA738672	vv64f10.r1 Stratagene	3.22e-50
c	43	295	15.8	362	10	AA292334	zt51b08.r1 Soares ovar	1.37e-49
c	44	294	15.7	478	20	AA862032	ol46e09.s1 NCI_CGAP_HN	2.84e-49
c	45	292	15.6	672	25	AI323721	mq40e08.x1 Barstead MP	1.21e-48

ALIGNMENTS

```
RESULT     1
LOCUS       AI557281    777 bp    mRNA         EST       23-MAR-1999
DEFINITION  PT2.1_15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION   AI557281
NID         g4489644
VERSION     AI557281.1  GI:4489644
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun
            Yu,J. and Hood,L.
  TITLE     Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187382.
            Contact: Guyang Matthew Huang
            Leroy Hood
            University of Washington
            Department of Molecular Biotechnology, Box 357730, University of
            Washington, Seattle, WA 98195
            Tel: 5106280100
            Fax: 5106280108
            Email: huangm@yahoo.com.
FEATURES             Location/Qualifiers
     source          1..777
                     /organism="Homo sapiens"
                     /note="Organ: Prostate; Vector: pBluescript; Directional
                     cDNA library was constructed using Lambda ZAP II kit
                     (Stratagene). mRNA was extracted from a frozen prostate
                     tumor tissue (Mayo Clinics)."
```

```
Insert Length: 640      Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 412.
```

FEATURES   Location/Qualifiers
     source   1..415
                   /organism="Homo sapiens"
                   /note="organ: prostate; Vector: pAMP10; mRNA made from
                   normal prostatic epithelial cells, cDNA made by oligo-dT
                   priming. Non-directionally cloned Size-selected on
                   agarose gel, average average insert size 600 bp. Library made by
                   D. Krizman, NIH."
                   /db_xref="taxon:9606"
                   /clone="IMAGE:996416"
                   /clone_lib="NCI_CGAP_Pr9"
                   /sex="male"
                   /tissue_type="normal prostatic epithelial cells"
                   /lab_host="DH10B"

BASE COUNT   78 a   125 c   129 g   83 t
ORIGIN

Query Match   74.4%;   Score 1392; DB 14;   Length 415;
Best Local Similarity   52.9%;   Pred. No. 0.00e+00;
Matches   203;   Conservative   99;   Mismatches   81;   Indels   1;   Gaps   1;

Db   33 GGTGGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTCGCTAACGA 92
           || ||:||| :: :| :: ||| |: | |: | | |:::|||: |:|||| : |:||: ||:||:
Qy   3 GGTGNARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGA   62

Db   93 CCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATCAG 152
           ::|| || ||| | || |: |||::| || ::|| : | :|||| |||: || :|:::|
Qy   63 YYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWS 122

Db   153 CATTGCTTCGCAGTGCCTACCGCCGGGGAACTCTTGCCTCGTTTCTGCCTCGTGGGTCTGCT 212
           ||| | :|| ||| | || | ||| |||| |: || | || | ||::| : -|| :|:
Qy   123 NATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTNYT 182

Db   213 GGCGAACGGCAGAATGC-TACCGTGCTGCAGTGCGTGAACGTGTCGCGGGTCTGTGTCTGAGGA 271
           || |: || | || | ||: || : || |:|| :|| : |: :|||| : :|:| |||:||
Qy   183 NGCNAAYGGNGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGA 242

Db   272 GGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGCGGAGG 331
           :| || : || |:| || : || | | | | |:::|| |: || :| :| ||:
Qy   243 RGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGNGG 302

Db   332 GCAAGACCAGAGGAGCTCCTGCAACGTGACTCTGGGGGCCCCTGATCTGCAACGGGTA 391
           ||: || |: | || |: | || |:||:|| || |||| : || || |:||||| |
Qy   303 NCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGNCNYTNATHTGYAAYGGNTA 362

Db   392 CTTGCAGGGCCTTGTGTCTTTCGG 415
           ::|| :|||:: || :|:||:|
Qy   363 YYTNCARGGNYTNGTNWSNTTYGG 386

RESULT   3
LOCUS   AA533140   576 bp   mRNA   EST   21-AUG-1997
DEFINITION nj46h06.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:995579
           similar to TR:G198491 G198491 KALLIKREIN ;contains MSR1.t1 MSR1
           repetitive element ;, mRNA sequence.
ACCESSION AA533140
NID   g2277236
VERSION   AA533140.1   GI:2277236
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 576)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Nov 29, 1993 this sequence version replaced gi:634881.
```

---

```
BASE COUNT    162 a    205 c    227 g    157 t    26 others
ORIGIN

Query Match              77.6%;    Score 1452;  DB 28;    Length 777;
Best Local Similarity    54.8%;    Pred. No. 0.00e+00;
Matches    256;    Conservative    107;    Mismatches    95;    Indels    9;    Gaps    9;

Db    185  ATGGTGGAGGCCAGCCTCCTCCGTACGGCACCCAGAGTACAACAGACCCTTGCTCGCTAAC  244
           ||||| |||||:: :| :: |||| || | |: | | |:::|||: |:|||| : |:||:  |
Qy      1  ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAY  60

Db    245  GACCTCATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTCTGACACCATCCGGAGCATC  304
           || :||| ||| |  || |: |||::| || ::|| : | :|||| |||: || :|::  ||:
Qy     61  GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH  120

Db    305  AGCATTGCTTCGCAGTGCCTACCGCCGGGGAACTCTTGCCTCGTTTCTGCCTCGTGGGTCTG  364
           :: :||| | :||  ||| | || | ||| |||| |: |  || |  ||::| : -|||  : ||
Qy    121  WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGNYTN  180

Db    365  CTGGCGAACGGCAGAATGC-TACCGTGCTGCAGTGCGTGTGCTGTGAG  423
           :: :|||| :: || | | ||: || : || |:|| :|| :||:|| :| |:
Qy    181  YTNGCNAAYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGAR  240

Db    424  GAGGTCTGCAGTAAGCTCTATGACCCGCTGTACCACCCCAGCATGTTCTGCGCCGGCGGA  483
           || :| || : || |:|  || :  || |  | | |  |:::|| |: || :| :| ||
Qy    241  GARGTNTGYWSNAARYTNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCNGGNGGN  300

Db    484  GGGCAAGACCAGAAGACTCCTGCAACGGTGACTCTGGGGGCC-CTGATNTGNAACGGG  542
           ||: || |: | || |: | || |:||:|| || ||||| : || || |:||||||
Qy    301  GGNCARNNNCARNNNGAYWSNTGYAAYGGNGNGGNCCNYTNATHTGYAAYGGN  360

Db    543  GNCTTGCAGGGCCTTGTCTTTCGGAAAAGC-CCGTGTGG-CAAGTTGGCCTGC-AGGT  599
           | :: :|||:: ||  :|:||:|  | | :  || |: | ||: ||| ||: :|:||
Qy    361  TAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGNCARGTNGGN  420

Db    600  GT-TACAC-AACCTCTG-AAATT-ACTGAGTGATAGANAAAACGGT  642
           || :|||  ||:: :|:| | | |  |:| |: || || ||||| |
Qy    421  GTNTAYACNAAYYTNTGYAARTTYACNGARTCGGATHGARAARACNGT  467

RESULT    2
LOCUS     AA551449          415 bp     mRNA          EST         05-SEP-1997
DEFINITION nj55e05.s1 NCI_CGAP_Pr9 Homo sapiens cDNA clone IMAGE:996416
           similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
           ;, mRNA sequence.
ACCESSION  AA551449
NID        g2321701
VERSION    AA551449.1    GI:2321701
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 415)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On Apr 14, 1993 this sequence version replaced gi:837437.

           Contact: Robert Strausberg, Ph.D.
           Tel: (301) 496-1550
           Email: Robert_Strausberg@nih.gov
           Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
           M.D., Michael R. Emmert-Buck, M.D., Ph.D.
           cDNA Library Preparation: David B. Krizman, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 885   Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 256.

FEATURES   Location/Qualifiers
     source   1..576
                     /organism="Homo sapiens"
                     /note"Organ: prostate; Vector: pAMP10; mRNA made from
                     normal prostatic epithelial cells, cDNA made by oligo-dT
                     priming. Non-directionally cloned. Size-selected on
                     agarose gel, average insert size 600 bp. Library made by
                     D. Krizman, NIH."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:995579"
                     /clone_lib="NCI_CGAP_Pr9"
                     /sex="male"
                     /tissue_type="normal prostatic epithelial cells"
                     /lab_host="DH10B"

BASE COUNT   169 a   156 c   139 g   102 t   10 others
ORIGIN

Query Match   32.7%;   Score 611; DB 14; Length 576;
Best Local Similarity 58.3%;   Pred. No. 1.65e-157;
Matches 98;   Conservative 36;   Mismatches 33;   Indels 1;   Gaps 1;

Db   45 CAGAAGGACTCCTGCAACGGTGACTCCTGTGGGGCCCCTGATCTGCAACGGGTACTTGCA 104
          ||:: ||:::|| ||::| |||| | || || ||:|| | :|||||||||:|:|||
Qy   310 CARNNNGAYWSNTGYAAYGNGAYWSNGGNGGN-CCNYTNATHTGYAAYGGNTAYYTNCA 368

Db   105 GGGCCTTGTGTCTTTCGGAAAAGCCCCGTGTGGCCAAGTTGGCGTGCCAGGTGTCTACAC 164
          :|| :| ||| :|||::|||||| ||::|:||||||| ||||:||||||||||||||
Qy   369 RGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNGTNTAYAC 428

Db   165 CAACCTCTGCAAATTCACTGAGTGGATAGAGAAAAACCGTCCAGGCCAG 212
          ||::||:||:|||:|| ||:|| || ||||||||| |||:|||| ::
Qy   429 NAAYYTNTGYAARTTYACNGARTGGATHGARAARACNGTNCARGCNWS 476

RESULT 4
LOCUS   AI557025   722 bp   mRNA   EST   23-MAR-1999
DEFINITION PT2.1_10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION   AI557025
NID   g4489388
VERSION   AI557025.1   GI:4489388
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 722)
AUTHORS   Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordo,D., Jun
            Yu,J. and Hood,L.
TITLE   Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL   Unpublished (1999)
COMMENT   On Jun 5, 1998 this sequence version replaced gi:3187123.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195

Tel: 510680100
Fax: 510680108
Email: huanggm@yahoo.com.

FEATURES   Location/Qualifiers
     source   1..722
                     /organism="Homo sapiens"
                     /note"Organ: Prostate; Vector: pBluescript; Directional
                     cDNA library was constructed using Lambda ZP II kit
                     (Stratagene). mRNA was extracted from a frozen prostate
                     tumor tissue (Mayo Clinics)."
                     /db_xref="taxon:9606"
                     /clone_lib="tumor2"

BASE COUNT   146 a   156 c   195 g   176 t   49 others
ORIGIN

Query Match   32.5%;   Score 608; DB 28; Length 722;
Best Local Similarity 48.1%;   Pred. No. 1.95e-156;
Matches 89;   Conservative 50;   Mismatches 45;   Indels 1;   Gaps 1;

Db   178 ATGGTGGAGGCCAGCCTCCTCCGTACGGNACCCAGNGTACAACAGACCCTTGCTCGCTAAC 237
          |||||||||::|:|| ::|||| ||: | :|:|||| :||:||||| || ||||| !!
Qy   1 ATGGTNARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAYAYMGNCCNYTNYTNGCNAAY 60

Db   238 GACCTCATGCTCATNAAGTNGGACGAATCCGTGTCCGAGTCTGANACCATCCGGAGCATC 297
          ||::|| ||| |||:|||:|:|| ||:|||||:|:||::|| :||:||||||||||| :|
Qy   61 GAYYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATH 120

Db   298 AGCATTGCTTCGNAGTGCCCTACCGTGGGGAACTCTTNCCTTCGTTTCTTCGCTGGGGTCT 357
          ::|:||: :| |:|||| || || || :|||: :::| :::| || ::|||||||||| :|
Qy   121 WSNATHGCNWSNCARTGYCCNACNCGNGGNAAYWSNTGYYTN-GTNWSNGGNTGGGGNYT 179

Db   358 GCTTG 362
          :| |
Qy   180 NYTNG 184

RESULT 5
LOCUS   AA503963   404 bp   mRNA   EST   20-AUG-1997
DEFINITION nh39a01.s1 NCI_CGAP_Pr5 Homo sapiens cDNA clone IMAGE:954696
            similar to SW:KLKA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
            ;contains MSR1.b2 MSR1 repetitive element ;, mRNA sequence.
ACCESSION   AA503963
NID   g2238930
VERSION   AA503963.1   GI:2238930
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 404)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:693367.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1244   Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 144.

FEATURES   Location/Qualifiers
     source   1..404

```
                  /organism="Homo sapiens"
                  /note="Vector: pAMP10; mRNA made from normal prostatic
                   epithelial cells, cDNA made by oligo-dT priming.
                   Non-directionally cloned. Size-selected on agarose gel,
                   average insert size 600 bp. "
                  /db_xref="taxon:9606"
                  /map="9"
                  /clone="IMAGE:954696"
                  /clone_lib="NCI_CGAP_Pr5"
                  /sex="male"
                  /tissue_type="prostate"
                  /lab_host="DH10B"
BASE COUNT      64 a     91 c    159 g     90 t
ORIGIN

Query Match       25.3%;  Score 474;  DB 14;  Length 404;
Best Local Similarity  58.9%;  Pred. No. 3.25e-109;
Matches   83;  Conservative 26;  Mismatches 31;  Indels 1;  Gaps 1;

Db  264 CTGGCCTGGACGGTTTCTCTATCCACTCAGGTGAATTTGCAGAGGTTGGTGTAGACACC 323
            ::: :: ::::: :::::::::::  :::::::::  ::::::  ::::: :::: ::
Cp  476 SWNGCYTGNACNGTYTNYTCDATCCAYTCNG-TRAAYTTRCANARRTTNGTRTANACNCC 418

Db  324 TGGCACGCCAACTTGGCCACAGGTGCTTTTCCGAGATACACAAGGCCCTGCAGGTACCC 383
            :: :: :: ::: :: ::: :::  ::::::  ::: : ::  :::: :: :: :::
Cp  417 NGGNACNCCNACYTGNCCNCANGGNGCYTTNCCRANGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCC 358

Db  384 GGTGCAGATTCAGGGGCCCCC 404
            :  ::::: :: :: :
Cp  357 RTTRCADATNARNGGNCCNCC 337

RESULT 6
LOCUS       AI385433     457 bp     mRNA            EST       27-JAN-1999
DEFINITION  mi85b05.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:473361 5' similar to gb:M22612 TRYPSINOGEN I PRECURSOR
            (HUMAN); gb:X04574 Mouse mRNA for preprotrypsin (MOUSE);, mRNA
            sequence.
ACCESSION   AI385433
NID         g4198896
VERSION     AI385433.1  GI:4198896
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 457)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston.R. and Wilson.R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     On May 1, 1998 this sequence version replaced gi:3105815.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:284105
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Putative full length read
            vector to vector length is 813
            Seq primer: -40RP from Gibco
            High quality sequence stop: 413.
FEATURES    Location/Qualifiers
```

```
source
              1..457
              /organism="Mus musculus"
              /note="Vector: pT7T3D (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
               was primed with a Not I - oligo(dT) primer [5'
               TGTTACCAATCGAAGTGGAGGCGGCGCATTTTTTTTTTTTTTTTTT 3'],
               double-stranded cDNA was size selected, ligated to Eco RI
               adapters (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of a modified pT7T3 vector
               (Pharmacia). Library went through one round of
               normalization to a Cot = 5. Library constructed by Bento
               Soares and M.Fatima Bonaldo. RNA was kindly provided by
               Dr. Minoru Ko (Wayne State University)."
              /db_xref="taxon:10090"
              /map="911H02; 2"
              /clone="IMAGE:473361"
              /clone_lib="Soares mouse p3NMF19.5"
              /dev_stage="19.5 dpc total fetus"
              /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      114 a    123 c    123 g     97 t
ORIGIN

Query Match       23.7%;  Score 443;  DB 26;  Length 457;
Best Local Similarity  38.4%;  Pred. No. 1.55e-98;
Matches  168;  Conservative 80;  Mismatches 180;  Indels 9;  Gaps 6;

Db   24 ATGTTCCAGGAATCAAATCATCCCCCACCCCGGTTACTCCCACCCTGGCCACTCCAAT 83
             :                   :  :  :       ::: :::: : ::    :
Qy    1 ATGGTNGARGCNWSNYTNWSNGTNMGNCAYCCNGARTAYAATMGNCCNYTNYTNGCNAAY 60

Db   84 GACCTCATGCTCATCAAAATGAACAGAAAAATCCGTGACTCCTCACTCAGTGAAGCCGTC 143
             ::: :: :  : :  : : :: :::  :  :       ::  : :: : : ::
Qy   61 GAYTNATGYTNATHAARYTNGAYGARWSNGTNWSNGAYACNATHMGNWSNATH 120

Db  144 GAAATTGCTTGTGACTGTGCCACCGAGGGACTAGGTGCATGGTGTCTGGCTGGGGGACA 203
             ::    : :   ::  :: : :   :: :::  : :::   :: ::  ::::: :
Qy  121 WSNATHGCNWSNCARTGYCCNACNGCNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGN-YT 179

Db  204 ACGAGCAGCAGCCACAATAACTTCCCGAAAGTCCTCCAGTGCCTGAATATTACTGTGCTC 263
             :  :          : ::  :: ::   : :: :: :::: ::: :: : : :: ::
Qy  180 NYTNGCNA-A--YGGNMGNA--TGCCNACNGTNYTNCARTGYGYTNAAYGTNWSNGTNGTN 234

Db  264 AGTGAGGAGAGGTGTAAAAACTCCTACCCAGGACAGATAGACAAGACCATGTTCTGCCCA 323
             :  ::::: :::       ::: :: :     ::        :  ::     ::: :
Qy  235 WSNGARGARGTNTGYWSNAARYTNNTAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYGCN 294

Db  324 GGTGATGAAGAGGG-CAGG--GACTCCTGCCAGGTGATTCCGAGGTCCTGTGGTCTGC 380
             :: ::    : ::   ::    : :: ::   :  : :: :::: : :: :: : :::
Qy  295 GGNGGNGGNCARNNNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGY 354

Db  381 AATGGCAAGTTACAGGGCCTTGTGTCCTGGGGTGATTTCCCCTGTGCTCAGCGGAACAGA 440
              ::::::::  : :: ::  :::: : :::: :: : ::::: ::  :: :::  :::
Qy  355 AAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTN 414

Db  441 CCAGGTGCTCACAACAA 457
            :: ::   : :: :::
Qy  415 CNNGGNGTNTAYACNAA 431

RESULT 7
LOCUS       W73168     498 bp     mRNA            EST       16-OCT-1996
DEFINITION  zd55e11.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:344588 5' similar to SW:TRY2_CANFA P06872 TRYPSINOGEN
            ANIONIC PRECURSOR ;, mRNA sequence.
ACCESSION   W73168
NID         g1383322
VERSION     W73168.1  GI:1383322
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 498)
```

```
DEFINITION   mb71h07.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
             IMAGE:334909 5' similar to gb:M22612 TRYPSINOGEN I PRECURSOR
             (HUMAN); gb:X04574 Mouse mRNA for preprotrypsin (MOUSE);, mRNA
             sequence.
ACCESSION    AI391329
NID          g4217336
VERSION      AI391329.1    GI:4217336
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 505)
AUTHORS      Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
             Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
             Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
             Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
             Waterston,R. and Wilson,R.
TITLE        The WashU-NCI Mouse EST Project 1999
JOURNAL      Unpublished (1999)
COMMENT      On Apr 14, 1993 this sequence version replaced gi:716991.

             Contact: Marra M/WashU-NCI Mouse EST Project 1999
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             This read is a RESEQUENCE of a previously sequenced mouse clone
             This read has been verified (found to hit its original self in the
             correct orientation)
             Seq primer:-40RP from Gibco
             High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1..505
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
                     /db_xref="taxon:10090"
                     /clone="IMAGE:334909"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT    128 a    131 c    136 g    110 t
ORIGIN
Query Match                       23.5%;  DB 26;  Length 505;
Best Local Similarity 43.2%;  Pred. No. 3.63e-97;
Matches  115;  Conservative   48;  Mismatches 100;   Indels   3;  Gaps    2;

Db      82  CCGAAAGTCCTCCAGTGCCTGAATATTACTGTGCTCCAGTGAGGAGAGGTGTAAAAACTCC 141
            || || || :| ||:|:| |  | :| | :|:|| ||::: :| ||
Qy     199  CCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTN 258

Db     142  TACCCAGGACAGATAGACAAGACCATGTTCTCGCCAGGTGATGAAGAGGG-CAGG--GAC 198
            || :: ||||||||||:| ||| :||:||:||:|:|| : || :: |  ||
Qy     259  TAYGAYCCNYTNTAYCAYCCNWSNATGTTYTGYCNGNGNGGNCARNNNCARNNNGAY 318

Db     199  TCCTGCCAGGTGATTCCGAGGTCCTGTGGTCTGCAATGGCAAGTCAAGTTACAGGCCTTGTG 258
            :: || :||| : |||:: || :| :|| ::||| :|| :|:||| :| :|| |:| ||
Qy     319  WSNTGYAAYGGNGAYWSNGGNGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTN 378
```

```
RESULT   8
LOCUS        AI391329        505 bp    mRNA              EST       02-FEB-1999
```

```
Db   259  TCCTGGGGGTGATTCCCCTGCTCAGCGGAACAGAGACCAGTGTCTACACCAACCTGTGT 318
            :: |::|:| |||:|| || ::|| | ||| || :|| ||||||:| |||:|| ||:|
Qy   379  WSNTTYGGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGNGTNTAYACNAAYYTNTGY 438

Db   319  GAGTTCGTTAAGTGGATTAAAGACAC 344
          |:||:  |:||||||:|:|| ||
Qy   439  AARTTYACNGARTGGATHGARAARAC 464


RESULT   9
LOCUS       AA293027          517 bp      mRNA           EST       12-AUG-1997
DEFINITION  zt54a12.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone.
            IMAGE:726142 5' similar to TR:G940540 G940540 SERINE PROTEASE
            HOMOLOGUE ;, mRNA sequence.
ACCESSION   AA293027
NID         g1940923
VERSION     AA293027.1  GI:1940923
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-Merck EST Project 1997
JOURNAL     Unpublished (1997)
COMMENT     On Jan 25, 1995 this sequence version replaced gi:637777.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 856    Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1..517
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCGCGGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="GDB:5938189"
                     /db_xref="taxon:9606"
                     /map="11 p15.3-p13"
                     /clone="IMAGE:726142"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT    112 a    152 c    136 g    117 t
ORIGIN

Query Match     22.9%;  Score 428;  DB 10;  Length 517;
Best Local Similarity   40.8%;  Pred. No. 2.07e-93;
Matches  109;  Conservative  44;  Mismatches  114;  Indels   0;  Gaps   0;

Db    27  TCCAGTGTGCATACATCCACCTGGTGTCCCGTGAGGAGTGTGAGCATGCTACCCTGCC 86
            | ||:||:|| |:|   ||| :| ::|| || |||||:|||:|:|| :: |:|  |
Qy   209  TNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNTGYWSNAARYTNTAYGAYCNY 268
```

```
Db    87  AGATCACCCAGAACATGTGTGTCTGGGGATGAGAAGTACGGAAGGATTCCTGCCAGG 146
          :|: ||||| || || | | || || ||:| ::|||:|||  : ||||| ||||||
Qy   269  TNTAYCAYCCNWSNATGTTYTGYGCNGNGNGNGGNCARNNNGAYWSNTGYAAYG 328

Db   147  GTGATTCGGGGGTCCCGTCGGTGTATGTGGAGACCACCTCCGAGGCCTTGTGTCATGGGTA 206
          |:|| :|:|||||| ||| :|| || ||:|| ||:| ||||:| :|:|| ||| ||||||
Qy   329  GNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNA 388

Db   207  ACATCCCTGGGATCAAAGGAGAGAGCCAGGAGTCTACACCAACCTCTGCAGATACACGA 266
           || | ||| |||:|  | |:| ::|| || ||:|||:: |: | ::| |: :|| :||
Qy   389  ARGCNCCNTGYGGNCARGTNGGNGTNCCNGNGTNTAYACNAAYYTNTGYAARTTYACNG 448

Db   267  ACTGGATCCAAAAACCATTCAGGCCA 293
          |||||||| ||||||| || |: |||:
Qy   449  ARTGGATHGARAARACNGTNCARGCNW 475


RESULT   10
LOCUS       AI415008          686 bp      mRNA           EST       09-FEB-1999
DEFINITION  mb71h07.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:334909 3' similar to gb:M22612 TRYPSINOGEN I PRECURSOR
            (HUMAN); gb:X04574 Mouse mRNA for preprotrypsin (MOUSE);, mRNA
            sequence.
ACCESSION   AI415008
NID         g4258512
VERSION     AI415008.1  GI:4258512
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 686)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished (1999)
COMMENT     On Apr 21, 1998 this sequence version replaced gi:3072910.

            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1..686
                     /organism="Mus musculus"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
                     /db_xref="taxon:10090"
                     /clone="IMAGE:334909"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT    153 a    164 c    184 g    185 t
ORIGIN
```

```
Query Match          22.4%;    Score 419;   DB 26;    Length 686;
Best Local Similarity 43.4%;    Pred. NO. 2.38e-90;
Matches  116;  Conservative  48;  Mismatches  99;  Indels  4;  Gaps  3;

Db  329  GTGTCTTTAATCCACTTAACGAACTCACACAGGTTGGTGTAGACACCTGGTCTGTTCCGC  388
              ||| :.: :||||||  :||:|| |:||:|  |||: | || ||
Cp  464  GTYTTYTCDATCCAVTCNGTRAAYTTRCANARRTTNGTRTANACNCCNGGNACNCCNACY  405

Db  389  TGAGCACAGGGAAATCACCCCAGGACACAAGGCCCTGAACTTGCCATTGCAGACCACA   448
              ||: ||| || ||  |:||: |:|||||: | :|||:|  ||||||||:|:|
Cp  404  TGNCCRCANGGNGCVTTNCCRAANSWNACNARNCCYTGNARRTANCCRTTRCADATNARN  345

Db  449  GGACCTCCGGAATCACCCTGGCAGGAGTCC--CTGCC-CTCTTCATCACCTGCGCAGAAC  505
              :|  |||||:| || :||  :|| ||| :|  ||||:  |||| ||:||||
Qy  276  YCCNWSNATGTTYTGYGCNGNGGNGGNCARNNNGAYWSNTGYAAYGGNGAYWS  335

Db  506  ATGGGTCCGCTGGTAGAAGTTTTTACACCTCTCCCTCACTGAGCACAGT  565
              ||  ||  ::|: .:   |  :||  |: |||| ||  || :||||
Cp  284  ATNS-WNGGRTGRTANARNGGRTCRTANARYTTNSWRCANACYTCYTCNSWNACNACNSW  226

Db  566  AATATTCAAGCACTCGAGGACTTTCGG  592
              : ||| .:||:|||.| ||  || ||
Cp  225  NACRTTNACRCAYTGNARNACNGTNGG  199


RESULT  11
LOCUS      AA411252      590 bp     mRNA        EST          08-AUG-1997
DEFINITION zt33b03.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
           IMAGE:724109 5' similar to TR:G940540 G940540 SERINE PROTEASE
           HOMOLOGUE ;, mRNA sequence.
ACCESSION  AA411252
NID        g2068793
VERSION    AA411252.1   GI:2068793
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 590)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
           Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
           Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
           Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE      WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT    On Sep 12, 1996 this sequence version replaced gi:1407458.

           Contact: Wilson RK
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Insert Length: 831    Std Error: 0.00
           Seq primer: -28m13 rev2 ET from Amersham
           High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1..590
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCCGGTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo."
                     /db_xref="GDB:935586"
                     /db_xref="taxon:9606"
```

```
                     /clone="IMAGE:724109"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT  127 a    186 c    141 g    136 t
ORIGIN

Query Match          21.8%;   Score 407;   DB 13;    Length 590;
Best Local Similarity 45.0%;   Pred. No. 2.76e-86;
Matches  90;  Conservative  33;  Mismatches  77;  Indels  0;  Gaps  0;

Db   22  CCAGAACATGTTGTGCTGCTGGGGATGAGAAGTACGGGAAGGATTCCTGCCAGGTGATTC  81
              :|  ||||||||| || :|      |:  ||||:::||  ::|||| || |:::::
Qy  276  YCCNWSNATGTTYTGYGCNGNGGNGGNCARNNNGAYWSNTGYAAYGGNGAYWS  335

Db   82  TGGGGTCCGCTGGTATGTGGAGACCACCTCCGAGGCCTTGTCTGTCATGGGTAACATCCC  141
              ||||| || :| ||:|: ||:| : :||:||| :|| |||| ||:|| || ||||
Qy  336  NGGNGGNCCNYTNATHTGYAAYGGNTAYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCC  395

Db  142  CTGTGGATCAAAGGAGAAGGCCAGGAGTCTACCACCAACGTCTGCAGATACACGAACTGGAT  201
              ||:|:   ||  | | |||:||:|:|| ||:||| |:||| |:|||||||||| || ||
Qy  396  NTGYGGNCARGTNGGNGTNCCNGGNGTNTAYACNAAYNTNGYAARTTYACNGARTGGAT  455

Db  202  CCAAAAACCATTCAGGCCA  221
              : |||:|| ||||||:||
Qy  456  HGARARARACNGTNCARGCNW  475


RESULT  12
LOCUS      W73140          586 bp      mRNA         EST         16-OCT-1996
DEFINITION zd5sell.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
           IMAGE:344588 3' similar to PIR:A53968 A53968 serine proteinase SCCE
           precursor - human ;, mRNA sequence.
ACCESSION  W73140
NID        g1383275
VERSION    W73140.1    GI:1383275
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 586)
AUTHORS    Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
           Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
           Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
           Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
           Wilson,R.
TITLE      The WashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT    On Apr 14, 1993 this sequence version replaced gi:693450.

           Contact: Wilson RK
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           Insert Length: 761    Std Error: 0.00
           Seq primer: ETprimer
           High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1..586
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCCGGTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
```

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W."
/db_xref="GDB:1269963"
/db_xref="taxon:9606"
/map="1q; 4q21.2-ter"
/clone="IMAGE:344588"
/clone_lib="Soares_fetal_heart_NbHH19w"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"

BASE COUNT   130 a   139 c   179 g   136 t   2 others
ORIGIN

Query Match   21.7%;   Score 406;   DB 34;   Length 586;
Best Local Similarity 46.6%;   Pred. No. 6.01e-86;
Matches   81;   Conservative   31;   Mismatches   62;   Indels 0;   Gaps 0;

```
Db  351 TGGCCTCGATGTGTTTCCTGGATCCACTTGGTGAACTTGCAGAGGTTCGTGTAGACACCCG 410
        :||:|: | |||| |:||:| :||||||| :||||:|||||::|||:||
Qy  475 WNGCYTGNACNGTYTTYCDATCCAYTCNGTRAAYTTRCANARRTTNGTRTANACNCCNG   416

Db  411 GTCTGTTGGGCCGGCACAAGGGTAATCTCCCAGGACACGAGTCCCTGCAGGAGCCAT    470
        : |:||| ||| | ||:| |||| :: ||: ||:||:| |:|| | |:|
Qy  415 GNACNCCNACYTGNCCRCANGGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCCRT 356

Db  471 TGCAGACCACAGGCCCCCCAGAATCACCCTGGCAAGAGTCTCTACTGCTTTGTC       524
        |:|:| | || ||:|| || :|| ||::|| |:||:||
Qy  355 TRCADAITNARNGGNCNCNCNSWRTCNCCRTTRCANSWRTCNNNYTGNNNYTGNC       302
```

RESULT 13
LOCUS   AA419284   625 bp   mRNA   EST   16-OCT-1997
DEFINITION zv35b02.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone
            IMAGE:755595 5' similar to gb:K01228 PROCOLLAGEN ALPHA 1(I) CHAIN
            PRECURSOR (HUMAN) ;, mRNA sequence.
ACCESSION   AA419284
NID   g2079014
VERSION   AA419284.1 GI:2079014
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 625)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,M., Lennon,G., Marra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE   WashU-Merck EST Project 1997
JOURNAL   Unpublished (1997)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1393686.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 290.
FEATURES   Location/Qualifiers
     source   1..625
                     /organism="Homo sapiens"
                     /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGACGCGGCCGCGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo. "
                     /db_xref="taxon:9606"
                     /clone="IMAGE:755595"
                     /clone_lib="Soares ovary tumor NbHOT"
                     /sex="Female"
                     /tissue_type="ovarian tumor"
                     /lab_host="DH10B (ampicillin resistant)"

BASE COUNT   157 a   183 c   157 g   128 t
ORIGIN

Query Match   20.4%;   Score 381;   DB 13;   Length 625;
Best Local Similarity 40.2%;   Pred. No. 1.48e-77;
Matches 107;   Conservative   42;   Mismatches 116;   Indels 1;   Gaps 1;

```
Db   90 ATGTGACCTTCCCTGACCTCATGTGCGTGGATGTCAAGCTCATCTCCCCCAGGACT     149
        |:|: |||:| | || || :| :||:||| |:|| :| ::| ||:||
Qy  188 AYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARGARGTNT  247

Db  150 GCACGAAGGTTACAAGGACTACTGGAAAATTCCATGCTGTCGCGTGGCATCCCCGACT   209
        || :| ||: | || :|||||:||| ::| ||||:| |||:|||:|| |:||
Qy  248 GYWSNAARYTNTAYGAYCCNYNTAYCAYCCNWSNAIGTTYTGYGCNGGNGNGNGNCARN   307

Db  210 CCAAGAAAAACGCCTGCAATGGTGACTCAGGGGGACCCTTGGTGTGCAGAGG-TACCTGC  268
        : |:| ||::||: | |||||: || |:| | |:|:| : || :
Qy  308 NNCARNNNGAYWSNTGYAAYGGNGAYWSNGGNGGNCCNYTNATHTGYAAYGGNTAYYTNC  367

Db  269 AAGGTCNGTGCTCCTGGGGAACTTCCCTGCGGCAACCCAATGACCCAGGAGTCTACA     328
        |:|:| | | :| ::|| |||| ||| | ::: || |::| :|| |||:||
Qy  368 ARGGNYTNGTNWSNTTYGNAARGCNCCNTGYGGNCARGTNGGNGTNCCNGGNGTNTAYA   427

Db  329 CTCAAGTGTGCAAAGTTCACCAGTGG                                   354
        || | | | ||| | :|| ||||||
Qy  428 CNAAYTYNTGYAARTTYACNGARTGG                                   453
```

RESULT 14
LOCUS   AA864127   496 bp   mRNA   EST   11-MAR-1998
DEFINITION vn42g07.r1 Stratagene mouse skin (#937311) Mus musculus cDNA clone
            IMAGE:1023900 5' similar to SW:SCCE_HUMAN P49862 STRATUM CORNEUM
            CHYMOTRYPTIC ENZYME PRECURSOR ;, mRNA sequence.
ACCESSION   AA864127
NID   g2956606
VERSION   AA864127.1 GI:2956606
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 496)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:716853.

            Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:574676
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 308.

```
FEATURES
     source
               Location/Qualifiers
               1..496
               /organism="Mus musculus"
               /strain="C57BL/6"
               /note="Organ: skin; Vector: pBluescript SK-; Site_1:
               EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
               Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
               Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
               adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
               sequence:5' CTCGAGTTTTTTTTTTTTTTTT 3'"
               /db_xref="taxon:10090"
               /clone="IMAGE:1023900"
               /clone_lib="Stratagene mouse skin (#937313)"
               /sex="females"
               /tissue_type="whole skin"
               /dev_stage="11 weeks old"
               /lab_host="SOLR (kanamycin resistant)"
BASE COUNT      125 a        134 c       137 g       100 t
ORIGIN

Query Match            20.3%;    Score 380; DB 20;    Length 496;
Best Local Similarity  35.2%;    Pred. NO. 3.20e-77;
Matches    163;  Conservative   81;  Mismatches 213;   Indels    6;   Gaps    3;

Db    33 AGGCTACAAAGTCATTCCGGCACCTGGCTACTCCACAAAGACCCACGTCAATGACATCA   92
         |:||  ::  :|:|  ||:||| ||: ::     |  ||  ::|| ||:||:||
Qy     8 ARGCNWSNYTNWSNGTNWGNCAYCCNGARTAYAAYMGNCCNYTNYTNGCNAAYGAYYTNA  67

Db    93 TGCTTGTGAGGTTGATGAGCCAGTCAAGATGTCATCCAAAGTGAGGCAGTCCAGCTTC  152
         |:|:| ||||| ||  |    ||  |:  ::    |     |:   :|   ::|
Qy    68 TGYTNATHAARYTNGAYGARWSNGTNWSNGARWSNGAYACNATHMGNWSNATHWSNATHG  127

Db   153 CTGAACACTGTGAACCCCCAGGGACGTCATGTACCGTCTCTGGATGGGCACCACAACCA  212
         |  ||  |||  ::    ||  : || :: ||  ||   || |||||
Qy   128 CNWSNCARTGYCCNACNGNGGNAAYWSNTGYYTNGTNWSNGGNTGGGGN---YTNYTN  183

Db   213 GCCCAGACGTGACCTTCCCTCGGATCTTATGTGCTCGGATGTGAAGCTCATCTCCTCCA  272
         || ||| |||  ::   ||  : |||: ||  ||::| : ::|:  :||:  :||:
Qy   184 GCN-A-AYGGNMGNATGCCNACNGTNYTNCARTGYGTNAAYGTNWSNGTNGTNWSNGARG  241

Db   273 GGGAGTGCAAGAAGGTGTACAAGAGACCTGCTGGGGAAAACCATGCTGTGTCTGGCATTC  332
         || |: || |||: :|| :| |:   |:|| | ::|:|
Qy   242 ARGTNTGYWSNAARYTNTAYGAYCCNYTNAYCAYCNWSNATGTTYTGYGCNGGNGGNG  301

Db   333 CTGACTCTAAGACAACAACAGTGCAATGGTGACTCAGGCGGGGCCCTGGTGTGCAACGACA  392
         ||  :| ||: :|:||| :||| || | |: ||  | |:|| ::|
Qy   302 GNCARNNNCARNNNGAYWSNTGYAAYGGNAYWSNGGNGGNCCNYTNATHTGYAAYGGNT  361

Db   393 CCCTCCAAGGTCTGGTGTCATGGGGTACGTACCCTTGTGGCCAGCACAACGAGCCAGGCG  452
         ::| ||||: |:|| :| ||| :: :     || |||   :  || |||
Qy   362 AYYTNCARGGNYTNGTNWSNTTYGGNAARGCNCCNTGYGCNCARGTNGGNGTNCCNGGNG  421

Db   453 TCTACACTCAAGTCTGCAAGTACTAACGGTGGGTGATGGAGAC  495
         |||  | | |: :::||:| :        :|  |||
Qy   422 TNTAYACNAAYYTNTGYAARTTYACNGARTGGATHGARARARAC  464
```

```
RESULT    15
LOCUS       AI139437        456 bp    mRNA        EST         27-OCT-1998
DEFINITION  qc20e03.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:1710172 3' similar to SW:SCCE_HUMAN P49862 STRATUM CORNEUM
            CHYMOTRYPTIC ENZYME PRECURSOR ;, mRNA sequence.
ACCESSION   AI139437
NID         g3645409
VERSION     AI139437.1   GI:3645409
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 456)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
```

```
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152039.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 637   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 424.

FEATURES
     source          Location/Qualifiers
                     1..456
                     /organism="Homo sapiens"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer (5'
                     TGTTACCAATCTGAAGTGGGAGCGGCGCATCTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHH19W."
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1710172"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
BASE COUNT       96 a      97 c      111 g      151 t      1 others
ORIGIN

Query Match            19.7%;   Score 368; DB 23;    Length 456;
Best Local Similarity  45.5%;   Pred. No. 3.13e-73;
Matches     77;  Conservative   27;  Mismatches 77;   Indels    0;   Gaps    0;

Db   253 TTTCATGGTGTCATTTATCCACTTGGTGAACTTGCACACTTGAGTGTAGACTCCTGGGTC  312
         :|   :   ||:|  ||:|| :|  |||||  ||:|  ::| |  || | :
Cp   471 YTGNACNGTYTYTCDATCCAYTCNGTRAAYTTRCANARRTTNGTRTANACNCNGNAC  412

Db   313 ATTGGGTTGGCCCAAGGAAGTTCCCCAGGACACCCAGACCTTGCAGGGTACCTCTGCA  372
         ::||||| ||  ::::   :| :|  |||  || |  |   || || ||
Cp   411 NCCNACYTGNCCRCANGGNGCYTTNCCRAANSWNACNARNCCYTGNARRTANCCRTTRCA  352

Db   373 CACCAACGGTCCCCCTGAGTCACCATTGCAGGCGTTTTCTTGGAGTGCGGGATGCCAGC  432
         ||  ||:  :|  :|  ||  | ||::| ||:::||  :|         ||||
Cp   351 DATNARNGGNCCNCNSWRTCNCCRTTRCANSWRTCNNNYTGNNNYTGNCNCCNCCNGC  292

Db   433 GCACAGCATGG  443
         :|| |||  |
Cp   291 RCARAACATNS  281
```

Search completed: Wed Sep 29 02:14:04 1999
Job time : 886 secs.

This Page Blank (uspto)

```
 __      __  ___    ___    ___    ___   __    __
|  \    /  ||   |  |   |  |   |  |   | |  |  |  |
|   \  /   ||   |  |   |  |   |  |   | |  |  |  |   (TM)
|    \/    ||   |  |   |  |   |  |   | |  |  |  |
|          ||   |  |   |  |   |  |   | |  |__|  |
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 28 10:55:54 1999;   MasPar time 8.69 seconds
                                              389.292 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172
Description: (1-159) from US09030606.pep
Perfect Score: 1145
Sequence: 1 MVEASLSVRHPEYNRPLLAN.........GVYTNLCKFTEWIEKTVQAS 159

Scoring table: PAM 150
                  Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database: a-geneseq35
                  1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
                  8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
                  14:part14 15:part15 16:part16 17:part17 18:part18
                  19:part19 20:part20 21:part21 22:part22 23:part23
                  24:part24 25:part25 26:part26 27:part27 28:part28
                  29:part29 30:part30 31:part31 32:part32 33:part33
                  34:part34 35:part35 36:part36 37:part37 38:part38
                  39:part39

Statistics: Mean 30.101;   Variance 116.165;   scale 0.259

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description	Pred. No.
1	1145	100.0	159	36	W71871	Protein encoded by pr	4.69e-111
2	1145	100.0	159	35	W69387	Prostate tumour speci	4.69e-111
3	1131	98.8	232	32	W59129	Homo sapiens Tub Inte	1.62e-109
4	1116	97.5	248	32	W60592	Human prostate-specif	7.22e-108
5	1114	97.3	205	35	W69388	Prostate tumour speci	1.20e-107
6	1114	97.3	205	36	W71872	Protein encoded by pr	1.20e-107
7	543	47.4	253	19	W05383	Human amyloid precurs	1.68e-45
8	543	47.4	253	13	R67888	Human stratum corneum	1.68e-45
9	506	44.2	244	23	W22985	Human serine protease	1.52e-41
10	506	44.2	244	9	R44532	Zyme APP-cleaving pro	1.52e-41
11	506	44.2	244	31	W51006	Protease M, a novel s	1.52e-41
12	495	43.2	246	35	W64260	Human amyloid beta-pr	2.27e-40
13	494	43.1	223	37	W81767	Bovine TRYP peptide f	2.91e-40
14	494	43.1	224	10	R53637	Bovine trypsin.	2.91e-40
15	494	43.1	230	10	R53638	Bovine trypsinogen.	2.91e-40
16	471	41.1	247	20	W08475	Porcine trypsinogen.	8.17e-38
17	469	41.0	260	20	W10694	Human recombinant neu	1.33e-37
18	469	41.0	260	21	W12393	Mouse neuropsin prote	1.33e-37
19	455	39.7	240	32	W57740	Trypsinogen-like prot	4.09e-36
20	447	39.0	260	39	W87703	A human serine protea	2.88e-35
21	438	38.3	164	35	W69389	Prostate tumour speci	2.59e-34
22	438	38.3	164	36	W71873	Protein encoded by pr	2.59e-34
23	420	36.7	247	1	P81243	Human spleen trypsin	2.07e-32
24	420	36.7	247	15	R82703	Human pancreatic tryp	2.07e-32
25	409	35.7	279	18	R94526	Korean Viper Salmosa	2.99e-31
26	406	35.5	276	20	W07620	Human NES1 polypeptid	6.20e-31
27	398	34.8	231	36	W76538	A. contortrix protein	4.31e-30
28	396	34.6	237	38	W83202	Prostate-specific gla	7.00e-30
29	396	34.6	237	30	W45395	Mature prostate-speci	7.00e-30
30	396	34.6	237	15	R84667	Mature kallikrein hK2	7.00e-30
31	396	34.6	244	30	W45396	Prostate-specific gla	7.00e-30
32	396	34.6	244	37	W83204	Prostate-specific gla	7.00e-30
33	396	34.6	261	38	W83203	Prostate-specific gla	7.00e-30
34	396	34.6	261	20	W06971	Prostate-specific gla	7.00e-30
35	396	34.6	261	20	W45397	Prostate-specific gla	7.00e-30
36	396	34.6	261	35	W49085	Wild-type human kalli	7.00e-30
37	393	34.3	234	4	R20557	Fibrinogenolytic prot	1.45e-29
38	392	34.2	237	37	W83212	hK2 variant A217v.	1.84e-29
39	392	34.2	237	35	W49087	Mutant human kallikre	1.84e-29
40	392	34.2	261	20	W45400	Prostate-specific gla	1.84e-29
41	392	34.2	261	20	W06972	Kallikrein prepro-hK2	1.84e-29
42	389	34.0	244	15	R84669	Pro-hK2 kallikrein.	3.81e-29
43	389	34.0	261	15	R84668	Prepro-hK2 kallikrein	3.81e-29
44	388	33.9	237	31	W56086	Human prostate specif	4.85e-29
45	388	33.9	261	26	W13649	Human prostatic speci	4.85e-29

ALIGNMENTS

```
RESULT   1
ID   W71871 standard; Protein; 159 AA.
AC   W71871;
DT   06-JAN-1999 (first entry)
DE   Protein encoded by prostate tumour clone P703 splice variant DE1.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
FH   Key              Location/Qualifiers
FT   Misc_difference  103
FT                    /note= "undefined residue"
FT   Misc_difference  105
FT                    /note= "undefined residue"
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03492.
PR   09-FEB-1998;  US-020956.
PR   25-FEB-1997;  US-806099.
PR   01-AUG-1997;  US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 3; Page 105; 130pp; English.
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence  159 AA;

Query Match             100.0%;  Score 1145;  DB 36;  Length 159;
Best Local Similarity   100.0%;  Pred. No. 4.69e-111;
Matches  159;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Db          1 mveaslsvrhpeynrpllandlmlikldesvsesdtirsisiasqcptagnsclvsgwgl 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy          1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60
```

Db　　61 langrmptvlqcvnvsvvseevcsklydplyhpsmfcagggqxqxdscngdsggplicng 120
　　　　　 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　61 LANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db　　121 ylqglvsfgkapcgqvgvpgvytnlckftewiektvqas 159
　　　　　 |||||||||||||||||||||||||||||||||||||||
Qy　　121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 2
ID　W69387 standard; Protein; 159 AA.
AC　W69387;
DT　08-DEC-1998　(first entry)
DE　Prostate tumour specific gene clone DEl protein.
KW　Prostate tumour specific gene; human; prostate cancer; detection;
KW　therapy.
OS　Homo sapiens.
FH　Key　　　　　Location/Qualifiers
FT　Misc_difference 103
FT　　　　　　/note= "unspecified amino acid"
FT　Misc_difference 105
FT　　　　　　/note= "unspecified amino acid"
PN　WO9837418-A2.
PD　27-AUG-1998.
PF　25-FEB-1998; U03690.
PR　09-FEB-1998; US-904809.
PR　25-FEB-1997; US-806596.
PR　01-AUG-1997; US-904809.
PA　(CORI-) CORIXA CORP.
PI　Dillon DC, Xu J;
DR　WPI; 98-480805/41.
DR　N-PSDB; V58644.
PT　Novel human prostate specific tumour protein and fragments - useful
PT　for detecting and treating prostate cancers
PS　Example 1; Page 112-113; 14lpp; English.
CC　This sequence is encoded by a human prostate tumour specific gene, and
CC　can be used in the method of the invention. The method is for detecting
CC　prostate cancer comprises contacting a biological sample with an agent
CC　able to bind an immunogenic portion of a prostate protein (such as
CC　this protein sequence). An antibody which binds to an immunogenic
CC　portion of the prostate protein, and the method can be used to detect,
CC　monitor progression of, or treat prostate cancers. The antibody may
CC　also be conjugated to a therapeutic agent for use in therapy of prostate
CC　cancers.
SQ　Sequence　159 AA;

Query Match　100.0%;
Best Local Similarity 100.0%;　Score 1145; DB 35; Length 159;
Matches　159;　Conservative　0;　Mismatches　0;　Indels　0;　Gaps　0;

Db　　1 mveaslsvrhpeynrpllandlmlikldesvsesdtirsisiasqcptagnsclvsgwgl 60
　　　　　 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60

Db　　61 langrmptvlqcvnvsvvseevcsklydplyhpsmfcagggqxqxdscngdsggplicng 120
　　　　　 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　61 LANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db　　121 ylqglvsfgkapcgqvgvpgvytnlckftewiektvqas 159
　　　　　 |||||||||||||||||||||||||||||||||||||||
Qy　　121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 3
ID　W59129 standard; Protein; 232 AA.
AC　W59129;
DT　11-SEP-1998　(first entry)
DE　Homo sapiens Tub Interactor (hTI-1) protein.
KW　serine protease; tub interactor; treatment; obesity; cachexia;
KW　anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW　neurodegenerative disease; Alzheimer's disease; drug screening;
KW　Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW　amyotrophic lateral sclerosis; spinocerebellar degeneration.

OS　Homo sapiens.
FH　Key　　　　　Location/Qualifiers
FT　Region　42
FT　　　　　　/note= "undefined amino acid"
PN　WO9812302-A1.
PD　26-MAR-1998.
PF　05-SEP-1997; U15627.
PR　21-JUL-1997; US-897340.
PR　17-SEP-1996; US-715032.
PA　(MILL-) MILLENNIUM PHARM INC.
PI　Errada PR, Gimeno CJ;
DR　WPI; 98-217246/19.
DR　N-PSDB; V11855.
PT　Tub interactor genes　- used to develop products for the treatment
PT　of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT　diabetes
PS　Claim 28; Fig 1; 120pp; English.
CC　The sequence is that encoding the Tub Interactor protein (hTI-1)
CC　which is a putative serine protease. TI genes function
CC　in biochemical pathways involved in weight control and
CC　related disorders. The products can be used for treating
CC　weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC　or a related disorder such as diabetes. The products can
CC　also be used to modulate cell cycle progression and apoptosis.
CC　They can be used for treating neurodegenerative diseases
CC　which are characterised by apoptosis, including Alzheimer's
CC　disease, Parkinson's disease, Huntington's chorea, amylotrophic
CC　lateral sclerosis or spinocerebellar degenerations. The
CC　products can also be used for detection, diagnosis and
CC　drug screening.
SQ　Sequence　232 AA;

Query Match　98.8%;　　Score 1131;　DB 32;　Length 232;
Best Local Similarity 97.5%;　Pred. No. 1.62e-109;
Matches　154;　Conservative　1;　Mismatches　3;　Indels　0;　Gaps　0;

Db　　66 mveaslsvrhpeynrpllandlmlikldesvsesdtirsisiasqcptagnsclvsgwgl 125
　　　　　 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy　　1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60

Db　　126 langrmptvlqcvnvsvvseevcsklydplyhpsmfcagggqdqtkdscngdsggplicng 185
　　　　　 ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Qy　　61 LANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db　　186 ylqglvsfgkapcgqvgvpgvytnlckftewiektvpg 223
　　　　　 |||||||||||||||||||||||||||||||||||||| :
Qy　　121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQA 158

RESULT 4
ID　W60592 standard; Protein; 248 AA.
AC　W60592;
DT　07-SEP-1998　(first entry)
DE　Human prostate-specific kallikrein (HPSK) protein.
KW　Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW　benign prostate hyperplasia; diagnosis; drug screening; PSK.
OS　Homo sapiens.
FH　Key　　　　　Location/Qualifiers
FT　Misc_difference 113
FT　　　　　　/label= unknown
FT　　　　　　/note= "encoded by NTC"
FT　Misc_difference 128
FT　　　　　　/label= unknown
FT　　　　　　/note= "encoded by AGN"
FT　Misc_difference 132
FT　　　　　　/label= unknown
FT　　　　　　/note= "encoded by GNT"
PN　WO9820117-A1.
PD　14-MAY-1998.
PF　31-OCT-1997; U20051.
PR　05-NOV-1996; US-744026.
PA　(INCY-) INCYTE PHARM INC.
PI　Bandman O, Goli SK;

```
DR   WPI; 98-286933/25.
DR   N-PSDB; V37495.
PT   New isolated prostate-specific kallikrein - used to develop products
PT   for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT   hyperplasia
PS   Claim 1; Fig 1A-C; 68pp; English.
CC   This represents a human prostate-specific kallikrein (HPSK). A host cell
CC   containing an expression vector comprising the HPSK nucleic acid sequence
CC   can be used to produce the protein recombinantly. The HPSK products can
CC   be used for the diagnosis of conditions or diseases associated with
CC   expression of HPSK such as prostate carcinoma and benign prostate
CC   hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC   modulate its activity can be used for the preparation of treatment of
CC   such conditions or diseases. The products can also be used for detection
CC   and drug screening, especially for the detection of prostate-specific
CC   kallikrein (PSK).
SQ   Sequence   248 AA;

Query Match             97.5%;    Score 1116;  DB 32;  Length 248;
Best Local Similarity 95.6%;    Pred. No. 7.22e-108;
Matches   152; Conservative    0; Mismatches    7;   Indels    0;  Gaps    0;

Db     90 mveaslsvrhpeynrpllandlmxikldesvsesdnirxisixsqcptagnfclvsgwgl 149
          ||||||||||||||||||| ||||||||| ||| ||| | || |||||| ||||||
Qy      1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60

Db    150 langrmptvlqcvnvsvvseevcsklydplyhpsmfcagggqdkdscngdsggplicng 209
          ||||||||||||||||| ||||||||||||||||||||| | |||||||||||||||
Qy     61 LANGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db    210 ylqglvsfgkapcgqvgvpgvytnlckftewiektvqas 248
          ||||||||||||||||||||||||||||||||||||||
Qy    121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 5
ID   W69388 standard; Protein; 205 AA.
AC   W69388;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone DE13 protein.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   Misc_difference 127
FT                   /note= "unspecified amino acid"
FT   Misc_difference 204
FT                   /note= "unspecified amino acid"
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   N-PSDB; V58647.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Example 1; Page 115-116; 141pp; English.
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence   205 AA;

Query Match             97.3%;    Score 1114;  DB 35;  Length 205;
```

```
Best Local Similarity 95.6%;    Pred. No. 1.20e-107;
Matches   152; Conservative    2; Mismatches    5;   Indels    0;  Gaps    0;

Db     47 mveaslsvrhpeynrllandlmlikldesvsesdtirsislasqcptagnsclvsgwgl 106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60

Db    107 langrmptvlhcvnvsvvsexvcsklydplyhpsmfcagggqdkdscngdsggplicng 166
          |||||||||:|||||| ||||||||||||||||||||||| | |||||||||||||||
Qy     61 LANGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db    167 ylqglvsfgkapcgqlgvpgvytnlckftewiektvqxs 205
          |||||||||||||:||||:|||||||||||||||||| |
Qy    121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 6
ID   W71872 standard; Protein; 205 AA.
AC   W71872;
DT   06-JAN-1999 (first entry)
DE   Protein encoded by prostate tumour clone P703 splice variant DE13.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   Misc_difference 127
FT                   /note= "undefined residue"
FT   Misc_difference 204
FT                   /note= "undefined residue"
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 3; Page 107-108; 130pp; English.
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence   205 AA;

Query Match             97.3%;    Score 1114;  DB 36;  Length 205;
Best Local Similarity 95.6%;    Pred. No. 1.20e-107;
Matches   152; Conservative    2; Mismatches    5;   Indels    0;  Gaps    0;

Db     47 mveaslsvrhpeynrllandlmlikldesvsesdtirsislasqcptagnsclvsgwgl 106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60

Db    107 langrmptvlhcvnvsvvsexvcsklydplyhpsmfcagggqdkdscngdsggplicng 166
          |||||||||:|||||| ||||||||||||||||||||||| | |||||||||||||||
Qy     61 LANGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db    167 ylqglvsfgkapcgqlgvpgvytnlckftewiektvqxs 205
          |||||||||||||:||||:|||||||||||||||||| |
Qy    121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 7
ID   W05383 standard; Protein; 253 AA.
AC   W05383;
DT   31-DEC-1996 (first entry)
DE   Human amyloid precursor protein protease.
KW   Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
KW   therapy.
OS   Homo sapiens.
```

```
Db    93 ikasksfrhpgystqthvndimlvklnsqarlssmvkkvrlpsrceppgttctvsgwgtt 152
         : | ||||| || ::::|||||:||::    ::.:|::| .:|||:: .:::|:|
Qy     2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db   153 tspdvtfpsdlmcvdvklispqdctkvykdllensmlcagipdskknacngdsgghplvcr 212
         :: .:| ||| : | ||| :| |||  ||||||||||:||||:
Qy    62 AN-G-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   213 gtlgglvswgtfpcgqpndpgvytqvckftkwindtm 249
         | ||||||| | |||| ::||| :||||| |:| |:
Qy   120 GYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTV 156
```

```
RESULT 9
ID    W22985  standard; Protein; 244 AA.
AC    W22985;
DT    08-OCT-1997  (first entry)
DE    Human serine protease 59 (SP59).
KW    Human; colon carcinoma; COLO 201; cell line; serine protease; SP59;
KW    screening; inhibitor; treatment; disease.
OS    Homo sapiens.
FH    Key              Location/Qualifiers
FT    peptide          1..21
FT                     /label= sig_peptide
FT    peptide          22..244
FT                     /label= mat_peptide
PN    J09149790-A.
PD    10-JUN-1997.
PF    24-JUL-1996; 212196.
PR    29-SEP-1995; JP-275105.
PA    (SUNR ) SUNTORY LTD.
DR    WPI; 97-357902/33.
DR    N-PSDB; T79126.
PT    Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
PT    - useful to screen for specific inhibitors, e.g. to search for, or
PT    study agent for treatment of various diseases.
PS    Claim 1; Pages 10-11; 16pp; Japanese.
CC    The present sequence is the human colon carcinoma COLO 201
CC    cell line derived serine protease 59 (SP59), which can be used to
CC    screen for specific inhibitors, e.g. to search for, or study an
CC    agent for the treatment of various diseases.
SQ    Sequence 244 AA;

Query Match             44.2%;    Score 506;   DB 23;   Length 244;
Best Local Similarity  42.1%;    Pred. No. 1.52e-41;
Matches    64; Conservative   33; Mismatches   55; Indels    0; Gaps    0;
```

```
Db    92 avihpdydaashdqdimllrlarpakiseliqplplerdcsanttschilgwgktadgdf 151
         :| :|||:: :|||::| ::||||||: :||:: :|| |||| ::|
Qy     7 SVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGRM 66

Db   152 pdtiqcayihlvsreecehaypgqitqnmlcagdekyykdscqgdsgplvcgdhlrglv 211
         :|:|| ||||: |||: : :|||:|| : |:| ||| :|:|| ||||::
Qy    67 PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLIV 126

Db   212 swnipcgskekpgvytvcrytnvwiqktiqa 243
         ||:|| :::|||:|| :|: |:||| |:
Qy   127 SFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQA 158
```

```
RESULT 10
ID    R44532  standard; Protein; 244 AA.
AC    R44532;
DT    01-JUL-1994  (first entry)
DE    Zyme APP-cleaving protease.
KW    Amyloid precursor protein-cleaving protease; Alzheimer's disease;
KW    Down's syndrome; diagnosis; propensity.
OS    Homo sapiens.
PN    EP-576152-A.
PD    29-DEC-1993.
PF    26-MAY-1993; 304103.
PR    28-MAY-1992; US-891542.
PA    (ElIL ) LILLY & CO ELI.
```

```
PN    WO9631122-A1.
PD    10-OCT-1996.
PF    02-APR-1996; U04294.
PR    04-APR-1995; US-416257.
PA    (ElIL ) LILLY & CO ELI.
PI    Dixon EP, Johnstone EM, Little SP;
DR    WPI; 96-464694/46.
DR    N-PSDB; T39783.
PT    New isolated human amyloid precursor protein protease - used to
PT    develop prods. for the treatment or diagnosis of associated
PT    conditions, esp. Alzheimer's disease
PS    Claim 1: Page 44-45; 55pp; English.
CC    Human amyloid precursor protein protease (W05383) is involved in
CC    the processing or clearance of amyloid precursor protein to form
CC    beta-amyloid peptide. Its amino acid sequence was deduced from
CC    a cDNA clone (T39783) obtd. from a human lung library. Recombinant
CC    protease can be produced in transformed or transfected prokaryotic
CC    (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC    used to develop products for the design and testing of cpds. useful
CC    for treating or preventing conditions associated with beta-amyloid
CC    peptide, esp. Alzheimer's disease.
SQ    Sequence 253 AA;

Query Match             47.4%;    Score 543;   DB 19;   Length 253;
Best Local Similarity  46.5%;    Pred. No. 1.68e-45;
Matches    73; Conservative   29; Mismatches   53; Indels    2; Gaps    2;
```

```
Db    93 ikasksfrhpgystqthvndimlvklnsqarlssmvkkvrlpsrceppgttctvsgwgtt 152
         : | ||||| || ::::|||||:||::    ::.:|::| .:|||:: .:::|:|
Qy     2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db   153 tspdvtfpsdlmcvdvklispqdctkvykdllensmlcagipdskknacngdsgghplvcr 212
         :: .:| ||| : | ||| :| |||  ||||||||||:||||:
Qy    62 AN-G-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   213 gtlgglvswgtfpcgqpndpgvytqvckftkwindtm 249
         | ||||||| | |||| ::||| :||||| |:| |:
Qy   120 GYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTV 156
```

```
RESULT 8
ID    R67888  standard; Protein; 253 AA.
AC    R67888;
DT    09-AUG-1995  (first entry)
DE    Human stratum corneum chymotrophic recombinant enzyme (SCCE).
KW    Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW    callosities; keratosis pilaris; ichthyoses; eczema.
OS    Homo sapiens.
PN    WO9500651-A.
PD    05-JAN-1995.
PF    20-JUN-1994; IB0166.
PR    18-JUN-1993; DK-000725.
PA    (SYMB-) SYMBICOM AB.
PI    Egelrud T, Hansson L;
DR    WPI; 95-052088/07.
DR    N-PSDB; Q81203.
PT    Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT    - and related vectors, transformed cells and polypeptides,
PT    useful for treating skin disorders, e.g. acne or psoriasis, and
PT    for identification of specific inhibitors.
PS    Disclosure; Page 97; 137pp; English.
CC    The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC    and skin care products, especially to treat and prevent acne,
CC    xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC    keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC    produced recombinantly following mammal, insect, plant, or
CC    microorganism transformation with plasmid pS507.
SQ    Sequence 253 AA;

Query Match             47.4%;    Score 543;   DB 13;   Length 253;
Best Local Similarity  46.5%;    Pred. No. 1.68e-45;
Matches    73; Conservative   29; Mismatches   53; Indels    2; Gaps    2;
```

PI   Dixon EP, Johnstone EM, Little SP, Norris FH;
DR   WPI; 94-001306/01.
DR   N-PSDB; Q53487.
PT   Amyloid precursor protein-cleaving protease - associated with
PT   Alzheimer's disease and Down's syndrome
PS   Claim 1; Page 11-12; 14pp; English.
CC   The sequence is that of zyme, an amyloid precursor protein-cleaving
CC   protease. It cleaves APP to generate amyloidogenic fragments of the
CC   size expected of a Met596-Asp597 cleavage. It is thus very useful
CC   in furthering the characterisation of Alzheimer's disease and
CC   Down's syndrome.
SQ   Sequence 244 AA;

Query Match   44.2%; Score 506; DB 9; Length 244;
Best Local Similarity 42.1%; Pred. No. 1.52e-41;
Matches   64; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

Db   92 avihpdydaashdqdimllrlarpakiseliqplplerdcsanttschilgwgktadgdf 151
          ::|:|: : ||:|| :||::::: :|: :|:|| :||
Qy   7 SVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGRM 66

Db   152 pdtiqcayihlvsreecehaypgqitqnmlcagdekygkdscqgdsggplvcgdhlrglv 211
          | : || : :|| | |||||||
Qy   67 PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLV 126

Db   212 swgnipcgskekpgvytnvcrytnwiqktiqa 243
          |: |: ||| |||||:|:|:|:||||
Qy   127 SFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA 158

RESULT 11
ID   W51006 standard; protein; 244 AA.
AC   W51006;
DT   04-AUG-1998 (first entry)
DE   Protease M, a novel serine protease.
KW   Protease M; serine protease; down-regulation; tumour; metastasis;
KW   antibody; antisense.
OS   Homo sapiens.
PN   WO9811238-A2.
PD   19-MAR-1998.
PF   11-SEP-1997; U16175.
PR   13-SEP-1996; US-025301.
PA   (DAND) DANA FARBER CANCER INST INC.
PI   Anisowicz A, Sager R, Sotiropoulou G;
DR   WPI; 98-207398/18.
DR   N-PSDB; V07152.
PT   DNA encoding a serine protease, designated protease M - used to
PT   identify modulators that can inhibit development or progression of
PT   metastatic phenotype(s) in mammary tumour cells
PS   Claim 28; Figure 2; 92pp; English.
CC   The invention relates to an isolated nucleic acid molecule encoding a
CC   novel serine protease, Protease M. Protease M is down-regulated in
CC   metastatic mammary epithelial tumour cells, as well as other tumour
CC   cells, and is up-regulated in senescent cells. The host cell is useful
CC   for the recombinant production of protease M. Probes and antibodies,
CC   based on the protease M gene and its protein product, are useful for
CC   detection of Protease M in a sample. Nucleic acids encoding protease M,
CC   antisense constructs, or antibodies against protease M are useful for
CC   modulating the protease M serine proteinase activity associated with a
CC   cell. In particular the modulators stimulate the protease M cysteine
CC   proteinase activity or inhibit the serine proteinase activity. The
CC   modulators are useful for inhibiting the development or progression of a
CC   metastatic phenotype in a (mammary) tumour cell. The modulators can be
CC   identified by methods measuring the rate of cleavage of protease M in
CC   the presence of a serine protease substrate and a test substance, or by
CC   measuring the level of expression of protease M mRNA in the presence of
CC   a test substance. The present sequence represents human protease M.
SQ   Sequence 244 AA;

Query Match   44.2%; Score 506; DB 31; Length 244;
Best Local Similarity 42.1%; Pred. No. 1.52e-41;
Matches   64; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

Db   92 avihpdydaashdqdimllrlarpakiseliqplplerdcsanttschilgwgktadgdf 151
          ::|:|: : ||:|| :||::::: :|: :|:|| :||
Qy   7 SVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGRM 66

Db   152 pdtiqcayihlvsreecehaypgqitqnmlcagdekygkdscqgdsggplvcgdhlrglv 211
          | : || : :|| | |||||||
Qy   67 PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLV 126

Db   212 swgnipcgskekpgvytnvcrytnwiqktiqa 243
          |: |: ||| |||||:|:|:|:||||
Qy   127 SFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA 158

RESULT 12
ID   W64260 standard; Protein; 246 AA.
AC   W64260;
DT   24-NOV-1998 (first entry)
DE   Human amyloid beta-protein precursor inhibitor.
KW   MCP-7; mast cell protease 7; tryptase-7; serine protease;
KW   amyloid beta-protein precursor inhibitor; human; blood clot;
KW   anticoagulant; myocardial infarction; reocclusion; thromboembolism;
KW   cerebral embolism; thrombosis; therapy.
OS   Homo sapiens.
PN   WO9824886-A1.
PD   11-JUN-1998.
PF   25-NOV-1997; U21620.
PR   04-DEC-1996; US-032354.
PA   (BGHM) BRIGHAM & WOMENS HOSPITAL.
PI   Stevens RL;
DR   WPI; 98-333308/29.
PT   New compositions containing tryptase-7, e.g. mouse mast cell
PT   protease7 - are used to treat clot formation in e.g. myocardial
PT   infarction, reocclusion following angioplasty or pulmonary
PT   thromboembolism
PS   Example; Page 77; 92pp; English.
CC   This polypeptide comprises the human amyloid beta-protein precursor
CC   inhibitor (ABPI). The crystallographic structure of the complex of
CC   rat trypsin with ABPI was used as a template structure to model the
CC   structure of the substrate-binding pocket of mouse mast cell
CC   protease 7 (mMCP-7, see W64233). The invention relates to mMCP-7
CC   and related tryptase-7 proteases that can be used to prevent or
CC   inhibit fibrin clot formation. Such proteases can be used to treat
CC   disorders mediated by undesirable thrombus clot formation such as
CC   myocardial infarction and reocclusion following angioplasty of
CC   blood clots associated with pulmonary thromboembolism, deep vein
CC   thrombosis, cerebral embolism, renal vein and peripheral arterial
CC   thrombosis.
SQ   Sequence 246 AA;

Query Match   43.2%; Score 495; DB 35; Length 246;
Best Local Similarity 43.1%; Pred. No. 2.27e-40;
Matches   69; Conservative 38; Mismatches 50; Indels 3; Gaps 3;

Db   88 inaakikhpnysswtlnndimliklsspvklnarvapvalpsacapagtqcllsgwgnt 147
          :| : ||:: : |||:|| |:|::::: : :::::|:| |:|||
Qy   2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db   148 lsngvnpdllqcvdapvlsqadceaaypgeitssmicvgfleggkdscqgdsggpvvcn 207
          |:| |||:|| : |||||::::|| ||||:||||
Qy   61 LANG-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   208 gqlqglvswgyg-calpdnpgvvytkvcnfvgwlqdtiaan 246
          | :|||:|:| | |:|:|||:|| |:|:|:|:||:::
Qy   120 GYIQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT 13
ID   W81767 standard; peptide; 223 AA.
AC   W81767;
DT   29-JAN-1999 (first entry)
DE   Bovine TRYP peptide fragment.
KW   DHFR; dihydrofolic acid reductase; protein function; trypsin; bovine;
KW   ribonuclease; myoglobin; database; homology; resemblance.

```
OS   Bos taurus.
PN   J10287696-A.
PD   27-OCT-1998.
PF   11-APR-1997; 093577.
PR   11-APR-1997; JP-093577.
PA   (IYAK-) IYAKU BUNSHI SEKKEI KENKYUSHO KK.
DR   WPI; 99-018384/02.
PT   Determination of protein biological function - comprises use of
PT   amino acid sequences database containing the relevant information
PS   Example 1; Fig 1; 11pp; Japanese.
CC   This sequence is used in the creation of a database containing the
CC   information for amino acid sequence of protein with at least 1
CC   biological function with added a score on importance of expression of
CC   the biological information for each amino acid residue. The database is
CC   useful for determination of unknown biological function of a protein or
CC   polypeptide based on the homology of amino acid sequence, e.g. steric
CC   structure of protein, and includes retrieval and evaluation of high
CC   homologous relationship for the determination of mostly resembling
CC   protein. The database allows for correct and rapid retrieval and
CC   presumption of protein and polypeptide having biological functions.
SQ   Sequence  223 AA;

Query Match             43.1%;  Score 494; DB 37; Length 223;
Best Local Similarity   43.1%;  Pred. No. 2.91e-40;
Matches   69; Conservative   36; Mismatches  52; Indels    3; Gaps    3;

Db   65 isasksivhpsynsntlndimiklksaaslnsrvasislptscasagtqclisgwgnt 124
         ||:||| | ||::|||| | ::|||| : :|||::: |:::| |::|| ||:|||
Qy    2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db  125 kssgtsypdvlkclkapilsdsscksaypgqitsnmfcagyleggkdscqgdsggpvvcs 184
         ::| ||| | |::.::|| :|||||| |||||||||||||::::|
Qy   62 -ANGRM-PTVLQCVNSVVSEEVCSKLIYDPLYHPSMFCAGGGQXDSCNGDSGGPLICN 119

Db  185 gklqgivswgsg-caqknkpgvytkvcnyvswikqtiasn 223
         | |||:||:| |-| |||:|:| || |||: || ::::
Qy  120 GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT    14
ID   R53637 standard; Protein; 224 AA.
AC   R53637;
DT   03-JAN-1995 (first entry)
DE   Bovine trypsin.
KW   Cattle; cow; trypsin; enzyme; protease; proinsulin; insulin;
KW   hormone; plasmid pRMG4.
OS   Bos taurus.
PN   EP-597681-A.
PD   18-MAY-1994.
PF   10-NOV-1993; 308959.
PR   13-NOV-1992; US-977703.
PA   (ELIL ) LILLY & CO ELI.
PI   Greaney MG, Rosteck PR;
DR   WPI; 94-160671/20.
DR   N-PSDB; Q63794.
PT   Expression vectors for bovine trypsin and bovine trypsinogen -
PT   for cleavage of zymogens into active drugs, e.g. pro-insulin
PT   conversion into insulin
PS   Disclosure; Page 25; 35pp; English.
CC   This protein may be expressed by E. coli/plasmid pRMG4 and is able
CC   to cleave zymogens into active drugs, e.g. pro-insulin conversion
CC   into insulin.
SQ   Sequence  224 AA;

Query Match             43.1%;  Score 494; DB 10; Length 224;
Best Local Similarity   43.1%;  Pred. No. 2.91e-40;
Matches   69; Conservative   36; Mismatches  52; Indels    3; Gaps    3;

Db   66 isasksivhpsynsntlndimiklksaaslnsrvasislptscasagtqclisgwgnt 125
         ||:||| | ||::|||| | ::|||| : :|||::: |:::| |::|| ||:|||
Qy    2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db  126 kssgtsypdvlkclkapilsdsscksaypgqitsnmfcagyleggkdscqgdsggpvvcs 185
```

```
Qy   62 -ANGRM-PTVLQCVNSVVSEEVCSKLIYDPLYHPSMFCAGGGQXDSCNGDSGGPLICN 119

Db  186 gklqgivswgsg-caqknkpgvytkvcnyvswikqtiasn 224
         | |||:||:| |-| |||:|:| || |||: || :|:
Qy  120 GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT    15
ID   R53638 standard; Protein; 230 AA.
AC   R53638;
DT   01-DEC-1994 (first entry)
DE   Bovine trypsinogen.
KW   Cattle; cow; trypsinogen; enzyme; protease; proinsulin; insulin;
KW   hormone; plasmid pRMG4.
OS   Bos taurus.
PN   EP-597681-A.
PD   18-MAY-1994.
PF   10-NOV-1993; 308959.
PR   13-NOV-1992; US-977703.
PA   (ELIL ) LILLY & CO ELI.
PI   Greaney MG, Rosteck PR;
DR   WPI; 94-160671/20.
PT   Expression vectors for bovine trypsin and bovine trypsinogen -
PT   for cleavage of zymogens into active drugs, e.g. pro-insulin
PT   conversion into insulin
PS   Disclosure; Page 27; 35pp; English.
CC   This protein may be expressed by E. coli/plasmid pRMG7 and is able
CC   to cleave zymogens into active drugs, e.g. pro-insulin conversion
CC   into insulin.
SQ   Sequence  230 AA;

Query Match             43.1%;  Score 494;  DB 10;  Length 230;
Best Local Similarity   43.4%;  Pred. No. 2.91e-40;
Matches   69; Conservative   52; Mismatches  35; Indels    3;  Gaps    3;

Db   72 isasksivhpsynsntlndimiklksaaslnsrvasislptscasagtqclisgwgnt 131
         ||:||| | ||::|||| | ::|||| : :|||::: |:::| |::|| ||:|||
Qy    2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db  132 kssgnsypdvlkclkapilsdsscksaypgqitsnmfcagyleggkdscqgdsggpvvcs 191
         ::| ||| | |::.::|| :|||||| |||||||||||||::::|
Qy   62 -ANGRM-PTVLQCVNSVVSEEVCSKLIYDPLYHPSMFCAGGGQXDSCNGDSGGPLICN 119

Db  192 gklqgivswgsg-caqknkpgvytkvcnyvswikqtias 229
         | |||:||:| |-| |||:|:| || |||: || ::
Qy  120 GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA 158

Search completed: Tue Sep 28 10:57:24 1999
Job time : 90 secs.
```

```
**********************************************

          W P S E A R C H (TM)

**********************************************

        Release 3.1A John F. Collins, Biocomputing Research Unit.
        Copyright (c) 1993-1998 University of Edinburgh, U.K.
                Distribution rights by Oxford Molecular Ltd.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 28 10:57:40 1999; MasPar time 9.35 Seconds
                                       681.280 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172
Description: (1-159) from US09030606.pep
Perfect Score: 1145
Sequence: 1 MVEASLSVRHPEYNRPLLAN........GVYTNLCKFTEWIERTVQAS 159

Scoring table: PAM 150
                Gap 11

Searched: 122810 seqs; 40068593 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database: pir60
                1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 41.036; Variance 70.441; scale 0.583

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	543	47.4	253	2	A53968	serine proteinase SCC	5.33e-98
2	519	45.3	248	2	S55066	trypsin (EC 3.4.21.4)	2.24e-92
3	516	45.1	247	2	A27547	trypsin (EC 3.4.21.4)	1.13e-91
4	514	44.9	246	2	JQ1472	trypsin (EC 3.4.21.4)	3.31e-91
5	512	44.7	246	2	B25528	trypsin (EC 3.4.21.4)	9.70e-91
6	505	44.1	246	2	JQ1471	trypsin (EC 3.4.21.4)	4.18e-89
7	504	44.0	248	2	S55067	trypsin (EC 3.4.21.4)	7.16e-89
8	503	43.9	248	2	S55065	trypsin (EC 3.4.21.4)	1.22e-88
9	500	43.7	247	2	S05494	trypsin (EC 3.4.21.4)	6.13e-88
10	497	43.4	246	1	TRDGC	trypsin (EC 3.4.21.4)	3.07e-87
11	495	43.2	246	1	TRRT1	trypsin (EC 3.4.21.4)	8.98e-87
12	495	43.2	247	1	TRRG	trypsin (EC 3.4.21.4)	8.98e-87
13	494	43.1	229	1	TRBOTR	trypsin (EC 3.4.21.4)	1.53e-86
14	494	43.1	250	2	TO1779	trypsin (EC 3.4.21.4)	1.53e-86
15	492	43.0	243	2	A35871	trypsin (EC 3.4.21.4)	4.49e-86
16	478	41.7	238	2	S31779	trypsin (EC 3.4.21.4)	8.09e-83
17	478	41.7	246	1	TRRT2	trypsin (EC 3.4.21.4)	8.09e-83
18	478	41.7	247	1	A25852	trypsin (EC 3.4.21.4)	8.09e-83
19	470	41.0	232	1	TRPGTR	trypsin (EC 3.4.21.4)	5.82e-81
20	469	41.0	260	2	I56559	neuropsin - mouse	9.92e-81
21	470	41.0	261	2	S01971	tissue kallikrein (EC	5.82e-81
22	468	40.9	247	2	S13813	trypsin (EC 3.4.21.4)	1.69e-80
23	464	40.5	261	2	JE0236	tissue kallikrein (EC	1.43e-79
24	463	40.4	242	2	S49489	trypsin (EC 3.4.21.4)	2.44e-79
25	462	40.3	261	2	A41020	tissue kallikrein (EC	4.15e-79
26	460	40.2	215	2	S66661	trypsin (EC 3.4.21.4)	1.21e-78
27	460	40.2	229	1	TRDFS	trypsin (EC 3.4.21.4)	1.21e-78
28	460	40.2	242	2	S31776	trypsin (EC 3.4.21.4)	1.21e-78
29	460	39.8	242	2	S31775	trypsin (EC 3.4.21.4)	1.21e-78
30	456	39.8	231	2	S31778	trypsin (EC 3.4.21.4)	1.02e-77
31	455	39.7	259	2	I38363	trypsin (EC 3.4.21.4)	1.73e-77
32	454	39.7	261	1	EGMSB	tissue kallikrein (EC	2.95e-77
33	455	39.7	304	2	S33396	trypsin (EC 3.4.21.4)	1.73e-77
34	453	39.6	453	2	B23863	tissue kallikrein (EC	5.02e-77
35	453	39.6	188	2	B32340	tissue kallikrein (EC	5.02e-77
36	452	39.5	149	1	KQMSM	tissue kallikrein (EC	8.54e-77
37	449	39.2	241	1	S39048	trypsin (EC 3.4.21.4)	4.21e-76
38	449	39.2	261	1	NGMSG	7S nerve growth facto	4.21e-76
39	446	39.0	261	2	A29745	tissue kallikrein (EC	2.08e-75
40	445	38.9	244	2	A44284	tissue kallikrein (EC	3.53e-75
41	444	38.8	247	1	B25852	trypsin (EC 3.4.21.4)	6.01e-75
42	440	38.4	261	2	A34079	tissue kallikrein (EC	5.03e-74
43	439	38.3	259	2	A29746	tissue kallikrein (EC	8.54e-74
44	438	38.3	259	2	B31136	tissue kallikrein (EC	1.45e-73
45	437	38.2	240	2	S39047	trypsin (EC 3.4.21.4)	2.47e-73

ALIGNMENTS

```
RESULT   1

ENTRY            A53968         #type complete
TITLE            serine proteinase SCCE precursor - human
ALTERNATE_NAMES  stratum corneum chymotryptic enzyme
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change
                 24-Sep-1998
ACCESSIONS       A53968
REFERENCE        A53968
   #authors      Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.;
                 Carlstein, A.; Egelrud, T.
   #journal      J. Biol. Chem. (1994) 269:19420-19426
   #title        Cloning, expression, and characterization of stratum corneum
                 chymotryptic enzyme. A skin-specific human serine
                 proteinase.
   #accession    A53968
     ##status    preliminary
     ##molecule_type mRNA
     ##residues  1-253 #label HAN
     ##cross-references GB:L33404; NID:g521214; PID:g532504
GENETICS
   #gene         GDB:PRSS6; SCCE
     ##cross-references GDB:377730
     ##map_position 7q35-7q35
CLASSIFICATION   #superfamily trypsin; trypsin homology
FEATURE
   30-245        #domain trypsin homology #label TRY
SUMMARY          #length 253   #molecular-weight 27525   #checksum 644

     Query Match             47.4%;   Score 543;   DB 2;   Length 253;
     Best Local Similarity   46.5%;   Pred. No. 5.33e-98;
     Matches   73;   Conservative   29;   Mismatches 53;   Indels  2;   Gaps  2;

Db    93 IKASKSFRHPGYSTQTHVNDLMLVKLNSQARLSSMVKKVRLPSRCEPPGTTCTVSGWGTT 152
            : :|| |||   |||||||| :||::   :: :::  :: :||||:|::|||||
Qy     2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISASQCPTAGNSCLVSGWGLL 61

Db   153 TSPDVTFPSDLMCVDVKLISPQDCTKVYKDLLENSMLCAGIPDSKKNACNGDSGGPLVCR 212
            :: ||||| : ||:: :|| :: ::|   : ||| ::::||:: :| ::|||||||| :||
Qy    62 AN-G-RMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   213 GTLQGLVSWGTFPCGQPNDPGVYTQVCKFTKWINDTM 249
            .|||||||| : |||  |:||||:|||||||| |::|: :
Qy   120 GYLQGLVSFGKAPCGVGVPCGVYTNLCKFTEWIEKTV 156

RESULT   2
```

```
ENTRY       S55066       #type complete
TITLE       trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-P29)
            - chicken
ALTERNATE_NAMES  trypsinogen II
ORGANISM    #formal_name Gallus gallus #common_name chicken
DATE        23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change
            08-Sep-1997
ACCESSIONS  S55066; S72347
REFERENCE   S55066; S55065
  #authors  Wang, K.; Gan, L.; Lee, I.; Hood, L.
  #journal  Biochem. J. (1995) 307:471-479
  #title    Isolation and characterization of the chicken trypsinogen
            gene family.
  #accession S55066
  ##molecule_type mRNA
  ##residues  1-248 ##label WAN1
  ##cross-references EMBL:U15157; NID:g603906; PID:g603906
  ##experimental_source clone 2-P29
  #accession S72347
  ##molecule_type DNA
  ##residues  1-248 ##label WAN2
  ##cross-references EMBL:U15157; NID:g603907; PID:g603907
  ##experimental_source clone 2-P29
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS    hydrolase; pancreas; protein digestion; serine proteinase;
            zymogen
FEATURE
   1-16     #domain signal sequence #status predicted #label SIG\
  17-25     #domain activation peptide #status predicted #label APT\
  26-248    #product trypsin II #status predicted #label MAT\
  26-241    #domain trypsin homology #label TRY\
  65,109,202 #active_site His, Asp, Ser #status predicted
SUMMARY     #length 248 #molecular-weight 26622 #checksum 1862

Query Match         45.3%;  Score 519;  DB 2;  Length 248;
Best Local Similarity 45.2%;  Pred. No. 2.24e-92;
Matches    71;  Conservative 37;  Mismatches 46;  Indels 3;  Gaps 3;

Db   92  SSVIIRHPKYSSITLNNDIMLIKLASAVEYSADIQPIALPSSCAKAGTECLISGWGNTLS 151
         ::|: :||| || ||||| ||||||| |:|||||| ::::||::| |: ||| ||||| :
Qy    4  ASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-LA 62

Db  152  NGYNYPELLQCLNAPILSDQECQEAYPGDITSNMICVGFLEGGKDSCQGDSGGPVVCNGE 211
         || |||| |:|::|: | |  |   |||:::|:| |||||||||||||||||:||||
Qy   63  NG-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGY 121

Db  212  LQGIVSWGIG-CALKGYPGVYTKVCNVVDWIQETIAA 247
         |||||:|| | ::| |||||||||:||||:|||: |
Qy  122  LQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA 158

RESULT 3
ENTRY       A27547       #type complete
TITLE       trypsin (EC 3.4.21.4) precursor, cationic - rat
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
            20-Mar-1998
ACCESSIONS  A27547
REFERENCE   A27547
  #authors  Fletcher, T.S.; Alhadeff, M.; Craik, C.S.; Largman, C.
  #journal  Biochemistry (1987) 26:3081-3086
  #title    Isolation and characterization of a cDNA encoding rat
            cationic trypsinogen.
  #cross-references MUID:87271609
  #accession A27547
  ##molecule_type mRNA
  ##residues  1-247 ##label FLE
  ##cross-references GB:M16624; NID:g206498; PID:g206499
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS    calcium binding; hydrolase; protein digestion; serine
            proteinase.
FEATURE
  25-240 ...  #domain trypsin homology #label TRY\
```

```
            31-161,49-65,
            133-234,140-207,    #disulfide_bonds #status predicted\
            172-186,            #active_site His, Asp, Ser #status predicted\
            64,108,201          #binding_site calcium (Glu, Asn, Val, Glu) #status
            76,78,81,86          predicted
SUMMARY     #length 247 #molecular-weight 26269 #checksum 9664

Query Match         45.1%;  Score 516;  DB 2;  Length 247;
Best Local Similarity 43.1%;  Pred. No. 1.13e-91;
Matches    69;  Conservative 41;  Mismatches 47;  Indels 3;  Gaps 3;

Db   89  IDAAKIIRHPSYNANTFDNDIMLIKLNSPATLNSRVSTVSLPRSCGSSGTKCLVSGWGNT 148
         ::|: :|||| || ||:::||||||| :: : :   ::  : :::|  |||:||||||
Qy    2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db  149  LSSGTNYPSLLQCLDAPVLSDSCKSYPGKITSNMFCLGFLEGGKDSCQGDSGGPVVCN 208
         |::| : :|||||:|:|::  : :||  |||:::| |||||||||||||::| ||||||
Qy   61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db  209  GQLQGVVSWGYG-CAQKGKPGVYTKVCNYVNWIQQTVAAN 247
         | |||| || | |:| | |||||||:|| ||:: :: ||
Qy  120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT 4
ENTRY       JQ1472       #type complete
TITLE       trypsin (EC 3.4.21.4) V precursor, b-form - rat
ORGANISM    #formal_name Rattus norvegicus #common_name Norway rat
DATE        17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change
            08-Sep-1997
ACCESSIONS  JQ1472
REFERENCE   JQ1471
  #authors  Kang, J.; Wiegand, U.; Mueller-Hill, B.
  #journal  Gene (1992) 110:181-187
  #title    Identification of cDNAs encoding two novel rat pancreatic
            serine proteases.
  #cross-references MUID:92165057
  #accession JQ1472
  ##molecule_type mRNA
  ##residues  1-246 ##label KAN
  ##cross-references EMBL:X59013; NID:g57414; PID:g57415
  ##experimental_source pancreas
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS    hydrolase; protein digestion; serine proteinase; zymogen
FEATURE
   1-15     #domain signal sequence #status predicted #label SIG\
  16-24     #domain activation peptide #status predicted #label ACT\
  25-246    #product trypsin V, b-form #status predicted #label MAT\
  25-239    #domain trypsin homology #label TRY\
            31-160,49-65,
            133-233,140-206,    #disulfide_bonds #status predicted\
            171-185,            #active_site His, Asp, Ser #status predicted
            64,108,200          #molecular-weight
SUMMARY     #length 246 #molecular-weight 26819 #checksum 4681

Query Match         44.9%;  Score 514;  DB 2;  Length 246;
Best Local Similarity 44.0%;  Pred. No. 3.31e-91;
Matches    70;  Conservative 39;  Mismatches 48;  Indels 2;  Gaps 2;

Db   89  IDAAKMILHPDYDKWTVDNDIMLIKLKSPATLNSKVSTIPLPQYCPTAGTECLVSGWGVL 148
         ::|: :||:|| || ||:::|||||| :: : :: ::  : :::|  ||| |||||||||
Qy    2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db  149  KFGFGESPSVLQCLDAPVLSDSVCHKAYPRQITNNMFCLGFLEGGKDSCQYDSGGPVVCNG 208
         |:| : :|||||:|:|::  : :|:  |: |||:::| |||||||||||| ::| ||||||
Qy   62  ANG-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db  209  EVQGIVSWGDG-CALEGKPGVYTKVCNYLNWIQQTVAAN 246
         | |||| || | |:| | |||||||:|| ||:: :: ||
Qy  121  YLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159
```

```
RESULT        5
ENTRY          B25528         #type complete
TITLE          trypsin (EC 3.4.21.4) precursor - mouse
ORGANISM       #formal_name Mus musculus #common_name house mouse mouse
DATE           30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change
               20-Mar-1998
ACCESSIONS     B25528
REFERENCE      A93646
  #authors     Stevenson, B.J.; Hagenbuechle, O.; Wellauer, P.K.
  #journal     Nucleic Acids Res. (1986) 14:8307-8330
  #title       Sequence organisation and transcriptional regulation of the
               mouse elastase II and trypsin genes.
  #cross-references MUID:87066713
  #accession   B25528
    ##molecule_type mRNA
    ##residues  1-246 ##label STE
    ##cross-references GB:X04574; NID:g54918; PID:g54919
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS       calcium binding; hydrolase; protein digestion; serine
               proteinase
FEATURE
  1-23         #domain signal sequence #status predicted #label SIG\
  24-246       #product trypsin #status predicted #label MAT\
  24-239       #domain trypsin homology #label TRY\
  30-160,48-64,
  132-233,139-206,   #disulfide_bonds #status predicted\
  171-185      #active_site His, Asp, Ser #status predicted\
  63,107,200   #binding_site calcium (Glu, Asn, Val, Glu) #status
  75,77,80,85             predicted
SUMMARY        #length 246 #molecular-weight 26203 #checksum 5833

Query Match     44.7%;  Score 512;  DB 2;  Length 246;
Best Local Similarity 43.9%;  Pred. No. 9.70e-91;
Matches  69;  Conservative  39;  Mismatches  46;  Indels  3;  Gaps  3;

Db     88 VDSAKIIRHPNYNSWTLDNDIMLIKLASPVTLNARVASVPLPSSCAPAGTQCLISGWGNT 147
          |::: :|||:|| | ||:|||| :|: .:|||  || :|::|:||| ||:||||
Qy      2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db    148 LSNGVNNPDLLQCVDAPVLPQADCEASYPGDITNNMICVGFLEGGKDSCQGDSGGPVVCN 207
          |:|| |||||:|| || ::::  |:| : ||| |:|  ||| ||||::| |||||||::||
Qy     61 LANG-RMPTVLQCVNVSVSSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db    208 GELQGIVSWGYG-CAQPDAPGVYTKVCNYVDWIQNTI 243
          ||::|::||| |:|  :||| ||||| |||::::|:
Qy    120 GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTV 156


RESULT        6
ENTRY          JQ1471         #type complete
TITLE          trypsin (EC 3.4.21.4) V precursor, a-form - rat
ORGANISM       #formal_name Rattus norvegicus #common_name Norway rat
DATE           17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change
               08-Sep-1997
ACCESSIONS     JQ1471; S23784
REFERENCE      JQ1471
  #authors     Kang, J.; Wiegand, U.; Mueller-Hill, B.
  #journal     Gene (1992) 110:181-187
  #title       Identification of cDNAs encoding two novel rat pancreatic
               serine proteases.
  #cross-references MUID:92165057
  #accession   JQ1471
    ##molecule_type mRNA
    ##residues  1-246 ##label KAN
    ##cross-references EMBL:X59012; NID:g57412; PID:g57413
    ##experimental_source pancreas
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS       hydrolase; protein digestion; serine proteinase; zymogen
FEATURE
  1-15         #domain signal sequence #status predicted #label SIG\
  16-24        #domain activation peptide #status predicted #label ACT\
  25-246       #product trypsin V, a-form #status predicted #label MAT\
```

```
  25-239       #domain trypsin homology #label TRY\
  31-160,49-65,
  133-233,140-206,   #disulfide_bonds #status predicted\
  171-185      #active_site His, Asp, Ser #status predicted
  64,108,200
SUMMARY        #length 246 #molecular-weight 26900 #checksum 4446

Query Match     44.1%;  Score 505;  DB 2;  Length 246;
Best Local Similarity 43.6%;  Pred. No. 4.18e-89;
Matches  68;  Conservative  38;  Mismatches  48;  Indels  2;  Gaps  2;

Db     89 IDAAKMILHPDYDKWTVDNDIMLIKLKSPATLNSKVSTIPLPQYCPFAGTECLVSGWGVL 148
          : |  |:|||||: ||:|||||||| :||.||:||||  :| |  |  ||:|||||||:
Qy      2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db    149 KFGFESPSVLQCLDAPVLSDSVCHKAYPRQITNNMFCLGFLEGGKDSCQYDSGGPVVCNG 208
          | :|||  ||| ||||||| :: |: |   :|| |  | ||| ||||::| |||||||:||
Qy     62 ANG-RMPTVLQCVNVSVSSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db    209 EVQGIVSWGDG-CALEGKPGVYTKVCNLNWIHQTI 243
          ||:||::|| : ||  |:|||||| |:  |::|:|:
Qy    121 YIQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTV 156


RESULT        7
ENTRY          S55067         #type complete
TITLE          trypsin (EC 3.4.21.4) I precursor, pancreatic (clone 1-P38) -
               chicken
ALTERNATE_NAMES trypsinogen I
ORGANISM       #formal_name Gallus gallus #common_name chicken
DATE           28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change
               08-Sep-1997
ACCESSIONS     S55067; S72345; S71155
REFERENCE      S55065
  #authors     Wang, K.; Gan, L.; Lee, I.; Hood, L.
  #journal     Biochem. J. (1995) 307:471-479
  #title       Isolation and characterization of the chicken trypsinogen
               gene family.
  #accession   S55067
    ##molecule_type mRNA
    ##residues  1-248 ##label WAN1
    ##cross-references EMBL:U15156
    ##experimental_source clone 1-P38
  #accession   S72345
    ##molecule_type DNA
    ##residues  1-248 ##label WAN2
    ##cross-references EMBL:U15156
    ##experimental_source clone 1-P38
REFERENCE      S71155
  #authors     Wang, K.
  #submission  submitted to the EMBL Data Library, September 1994
  #accession   S71155
    ##molecule_type mRNA
    ##residues  1-102,'A',104-248 ##label WAW
    ##cross-references EMBL:U15156; NID:g603904; PID:g603905
    ##experimental_source clone 1-P38
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS       hydrolase; pancreas; protein digestion; serine proteinase;
               zymogen
FEATURE
  1-15         #domain signal sequence #status predicted #label SIG\
  16-25        #domain activation peptide #status predicted #label APT\
  26-248       #product trypsin I #status predicted #label MAT\
  26-241       #domain trypsin homology #label TRY\
  65,109,202   #active_site His, Asp, Ser #status predicted
SUMMARY        #length 248 #molecular-weight 26103 #checksum 6242

Query Match     44.0%;  Score 504;  DB 2;  Length 248;
Best Local Similarity 43.1%;  Pred. No. 7.16e-89;
Matches  69;  Conservative  38;  Mismatches  50;  Indels  3;  Gaps  3;

Db     90 ISSSKVIRHSGYNSNTLNNDIMLIKLSKAATLNSYVNTVPLFTSCVTAGTTCLISGWGNT 149
          : :||: || || | ||:||||||| :|  || |:: | | :|  ||:|| |||||||
```

```
Qy      2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSEGDTIRSISIASQCPTAGNSCLVSGWG-L  60
              |::| :| :-||||:-:|:|    |       |      |||  ||||||:|:|
Db    150  LSSGSLYPDVLQCLNAPVLSSSQCSSAYPGRITSNMICIGYLNGGKDSCQGDSGGPVVCN  209
              |::| : - |||||-:|:|   |   |        |||  |:|||||::||
Qy     61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXXDSCNGDSGGPLICN  119

Db    210  GQLQGFVSWGIG-CAQKGYPGVYTKVCNYVSWIKTTMSSN  248
              | |||:|| |-|    ||::||  || |:|:|||||
Qy    120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS  159

RESULT       8
ENTRY        S55065      #type complete
TITLE        trypsin (EC 3.4.21.4) I precursor, pancreatic (clone P1) -
             chicken
ORGANISM     #formal_name Gallus gallus #common_name chicken
DATE         23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change
             08-Sep-1997
ACCESSIONS   S55065; S72346
REFERENCE    S55065
#authors     Wang, K.; Gan, L.; Lee, I.; Hood, L.
#journal     Biochem. J. (1995) 307:471-479
#title       Isolation and characterization of the chicken trypsinogen
             gene family.
#accession   S55065
##molecule_type mRNA
##residues   1-248  ##label WAN1
##cross-references EMBL:U15155; NID:g603902; PID:g603902
#experimental_source clone P1
#accession   S72346
##molecule_type DNA
##residues   1-248  ##label WAN2
##cross-references EMBL:U15155; NID:g603902; PID:g603903
#experimental_source clone P1
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS     hydrolase; pancreas; protein digestion; serine proteinase;
             zymogen
FEATURE
   1-15      #domain signal sequence #status predicted #label SIG\
   16-25     #domain activation peptide #status predicted #label APT\
   26-248    #product trypsin I #status predicted #label MAT\
   26-241    #domain trypsin homology #label TRY\
   65,109,202 #active_site His, Asp, Ser #status predicted
SUMMARY      #length 248  #molecular-weight 26069 #checksum 5893

Query Match  43.9%;  Score 503;  DB 2;  Length 248;
Best Local Similarity 43.1%;  Pred No. 1.22e-88;
Matches  69;  Conservative  38;  Mismatches  50;  Indels  3;  Gaps  3;

Db     90  ISSSKVIRHSGYNANTLNNDIMLIKLSKAATLNSYVNTVPLPTSCVTAGTTCLISGWGNT  149
              ::|| :||| || |||:|||||   |  ::-||  ||  |||  ||:||||
Qy      2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSEGDTIRSISIASQCPTAGNSCLVSGWG-L  60
              |::| :| :- |||||-:|:|    |       |      |||  ||||||:|:|
Db    150  LSSGSLYPDVLQCLNAPVLSSSQCSSAYPGRITSNMICIGYLNGGKDSCQGDSGGPVVCN  209
              |::| : - |||||-:|:|   |   |        |||  |:|||||::||
Qy     61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXXDSCNGDSGGPLICN  119

Db    210  GQLQGFVSWGIG-CAQKGYPGVYTKVCNYVSWIKTTMSSN  248
              | |||:|| |-|    ||::||  || |:|:|||||
Qy    120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS  159

RESULT       9
ENTRY        S05494      #type complete
TITLE        trypsin (EC 3.4.21.4) IV precursor - rat
ALTERNATE_NAMES 23K protein; trypsinogen IV precursor
ORGANISM     #formal_name Rattus norvegicus #common_name Norway rat
DATE         07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change
             08-Sep-1997
ACCESSIONS   S05494
REFERENCE    S05494
#authors     Luetcke, H.; Rausch, U.; Vasiloudes, P.; Scheele, G.A.; Kern,
             H.F.
```

```
#journal     Nucleic Acids Res. (1989) 17:6736
#title       A fourth trypsinogen (P23) in the rat pancreas induced by
             CCK.
#cross-references MUID:89386010
#accession   S05494
##molecule_type mRNA
##residues   1-247  ##label LUE
##cross-references EMBL:X15679; NID:g56813; PID:g56814
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS     calcium binding; hydrolase; protein digestion; serine
             proteinase; zymogen
FEATURE
   1-15      #domain signal sequence #status predicted #label SIG\
   16-23     #domain activation peptide #status predicted #label APT\
   24-247    #product trypsin IV #status predicted #label MAT\
   24-240    #domain trypsin homology #label TRY\
   30-161,49-65,
   133-234,140-207,
   172-186           #disulfide_bonds #status predicted\
   64,108,201        #active_site His, Asp, Ser #status predicted\
   76,78,81,86       #binding_site calcium (Glu, Asn, Val, Glu) #status
                     predicted
SUMMARY      #length 247  #molecular-weight 26573  #checksum 6086

Query Match  43.7%;   Score 500;  DB 2;  Length 247;
Best Local Similarity 42.0%;   Pred No. 6.13e-88;
Matches  66;  Conservative  39;  Mismatches  49;  Indels  3;  Gaps  3;

Db     89  IDAEKIIRHPEYNKDTLDNDIMLIKLKSPAVLNSQVSTVSLPRSCASTDAQCLVSGWGNT  148
              ::| :||||:: |||||:   |        ||||||
Qy      2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSEGDTIRSISIASQCPTAGNSCLVSGWG-L  60

Db    149  VSIGGKYPALLQCLEAPVLSASSCKKSYPGQITSNMFCLGFLEGGKDSCCGDSGGPVVCN  208
              :: |:| |::|||::|:|    |  |       |||||||:|:|||||::||:
Qy     61  LA-NGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN  119

Db    209  GEIQGIVSWGSV-CAMRGKPGVYTKVCNYLSWIQETM  244
              | |:||:||    |   |:|||  || |:|:|  ||
Qy    120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTV  156

RESULT       10
ENTRY        TRDGC      #type complete
TITLE        trypsin (EC 3.4.21.4) precursor, cationic - dog
ALTERNATE_NAMES cationic trypsinogen
ORGANISM     #formal_name Canis lupus familiaris #common_name dog
DATE         30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
             22-Jan-1999
ACCESSIONS   B26273
REFERENCE    A26273
#authors     Pinsky, S.D.; LaForge, K.S.; Scheele, G.
#journal     Mol. Cell. Biol. (1985) 5:2669-2676
#title       Differential regulation of trypsinogen mRNA translation:
             full-length mRNA sequences encoding two oppositely charged
             trypsinogen isoenzymes in the dog pancreas.
#cross-references MUID:86284628
#accession   B26273
##molecule_type mRNA
##residues   1-246  ##label PIN
##cross-references GB:M11590; NID:g164096; PID:g164097
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS     hydrolase; pancreas; protein digestion; serine proteinase;
             zymogen
FEATURE
   1-15      #domain signal sequence #status predicted #label SIG\
   16-23     #domain activation peptide #status predicted #label APT\
   24-246    #product trypsin, cationic #status predicted #label ENZ\
   24-239    #domain trypsin homology #label TRY\
   30-160,48-64,
   132-233,139-206,
   171-185           #disulfide_bonds #status predicted\
   63,107,200        #active_site His, Asp, Ser #status predicted\
   75,77,80,85       #binding_site calcium (Glu, Asn, Val, Glu) #status
```

```
SUMMARY              #length 246 #molecular-weight 26170 #checksum 7929
                     predicted

Query Match          43.4%;   Score 497;  DB 1;  Length 246;
Best Local Similarity 42.5%;  Pred. No. 3.07e-87;
Matches    68;  Conservative   38;  Mismatches   51;  Indels    3;  Gaps    3;

Db      88  INAAKIIRHPRYNANTIDNDIMLIKLSSPATLNSRVSAIALPKSCPAAGTQCLISGWGNT 147
            ::!:  :!!!!!!  !!:!!!!!!  : :::!!!!!!!  !!:!!  !!:!!!!
Qy       2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db     148  QSIQQNYPDVLQCLKAPILSDSVCRNAYPGQISSNMMCLGYMEGGKDSCQGDSGGPVVCN 207
            : !!:!:  ::!!::!! :  !!!! !!  ::!!::  ::  !!!  !!! !!!!!!!!!!!
Qy      61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGQOXDSCNGDSGGPLICN 119

Db     208  GELQGIVSWG-AGCAQKGKPGVSPKVCKVVSWIQQTIAAN 246
            !!!!!!:!! ! !! :  !!!!  :!! !! :! :!
Qy     120  GYLQGLVSFGKAPCGQVGPGVVPGVYTNLCKFTEWIEKTVQAS 159

RESULT  11
ENTRY      TRRT1       #type complete
TITLE      trypsin (EC 3.4.21.4) I precursor - rat
ALTERNATE_NAMES  trypsinogen I
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change
           20-Mar-1998
ACCESSIONS B22657; A00948
REFERENCE  A22657
#authors   Craik, C.S.; Choo, Q.L.; Swift, G.H.; Quinto, C.; MacDonald,
           R.J.; Rutter, W.J.
#journal   J. Biol. Chem. (1984) 259:14255-14264
#title     Structure of two related rat pancreatic trypsin genes.
#cross-references MUID:85054880
#accession B22657
##molecule_type DNA
##residues 1-246 ##label CRA
##cross-references GB:J00778; NID:g206507; PID:g206508
#note      the authors translated the codon ATC for residue 6 as
           Leu and GAC for residue 170 as Asn
REFERENCE  A00948
#authors   MacDonald, R.J.; Stary, S.J.; Swift, G.H.
#journal   J. Biol. Chem. (1982) 257:9724-9732
#title     Two similar but nonallelic rat pancreatic trypsinogens.
           Nucleotide sequences of the cloned cDNAs.
#cross-references MUID:82265624
#accession A00948
##molecule_type mRNA
##residues 1-246 ##label MAC
GENETICS
#introns   14/1; 67/2; 152/1; 197/3
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS   hydrolase; pancreas; protein digestion; serine proteinase;
           zymogen
FEATURE
  1-15     #domain signal sequence #status predicted #label SIG\
 16-23     #domain activation peptide #status predicted #label APT\
 24-246    #product trypsin I #status predicted #label ENZ\
 24-239    #domain trypsin homology #label TRY\
 30-160,48-64,
132-233,139-206,
171-185    #disulfide_bonds #status predicted\
 63,107,200 #active_site His, Asp, Ser #status predicted\
 75,77,80,85 #binding_site calcium (Glu, Asn, Val, Glu) #status
           predicted
SUMMARY    #length 246 #molecular-weight 25959 #checksum 6732
           predicted

Query Match          43.2%;   Score 495;  DB 1;  Length 246;
Best Local Similarity 43.1%;  Pred. No. 8.98e-87;
Matches    69;  Conservative   38;  Mismatches   50;  Indels    3;  Gaps    3;

Db      88  INAAKIIKHPNYSSWTLNNDIMLIKLSSPVKLNARVAPVALPSACAPAGTQCLISGWGNT 147
            ::!:  :!!!:: | !!:!!!!!  : :::!! | !: !!! !!:!!!!!
```

```
Qy       2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db     148  LSNGVNNPDLLQCVDAPVLSQADCEAAYPGEITSSMICVGFLEGGKDSCQGDSGGPVVCN 207
            : !!:!  ::!: :  :!:: : ! !!!!!!:!!!!!!!!!:::!::
Qy      61  LANG-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGQOXDSCNGDSGGPLICN 119

Db     208  GQLQGIVSWGYG-CALPDNPGVYTKVCNFVGVWIQQDTIAAN 246
            | !!!!!:!! ! !!! !!: :! !!!!!! !! :! !!
Qy     120  GYLQGLVSFGKAPCGQVGPGVVPGVYTNLCKFTEWIEKTVQAS 159

RESULT  12
ENTRY      TRDG        #type complete
TITLE      trypsin (EC 3.4.21.4) precursor, anionic - dog
ALTERNATE_NAMES  cationic trypsinogen
ORGANISM   #formal_name Canis lupus familiaris #common_name dog
DATE       30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
           22-Jan-1999
ACCESSIONS A26273
REFERENCE  A26273
#authors   Pinsky, S.D.; LaForge, K.S.; Scheele, G.
#journal   Mol. Cell. Biol. (1985) 5:2669-2676
#title     Differential regulation of trypsinogen mRNA translation:
           full-length mRNA sequences encoding two oppositely charged
           trypsinogen isoenzymes in the dog pancreas.
#cross-references MUID:86284628
#accession A26273
##molecule_type mRNA
##residues 1-247 ##label PIN
##cross-references GB:M11589; NID:g164094; PID:g164095
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS   hydrolase; pancreas; protein digestion; serine proteinase;
           zymogen
FEATURE
  1-15     #domain signal sequence #status predicted #label SIG\
 16-23     #domain activation peptide #status predicted #label APT\
 24-247    #product trypsin, anionic #status predicted #label ENZ\
 24-239    #domain trypsin homology #label TRY\
 30-160,48-64,
132-233,139-206,
171-185    #disulfide_bonds #status predicted\
 63,107,200 #active_site His, Asp, Ser #status predicted\
 75,77,80,85 #binding_site calcium (Glu, Asn, Val, Glu) #status
           predicted
SUMMARY    #length 247 #molecular-weight 26623 #checksum 8431

Query Match          43.2%;   Score 495;  DB 1;  Length 247;
Best Local Similarity 41.3%;  Pred. No. 8.98e-87;
Matches    66;  Conservative   43;  Mismatches   48;  Indels    3;  Gaps    3;

Db      88  INSAKVIRHPNYNSWILDNDIMLIKLSSPAVLNARVATISLPRACAAPGTQCLISGWGNT 147
            :::: :!!!!! :!:!!!!  :!!:: !!!:!!!   || :!!!!
Qy       2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db     148  LSSGTNYPELLQCLDAPILTQAQCEASYPGQITENMICAGFLEGGKDSCQGDSGGPVVCN 207
            | !!:!  ::!! !! :  :! :!: !!!!!! !!!!!!!!!!!
Qy      61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGQOXDSCNGDSGGPLICN 119

Db     208  GELQGIVSWYG-CAQKNKPGVYTKVCNFVDWIQQTIAAN 246
            !!!!!:!! ! !!:  !! :!! !!!!!! !! :! !!
Qy     120  GYLQGLVSFGKAPCGQVGPGVVPGVYTNLCKFTEWIEKTVQAS 159

RESULT  13
ENTRY      TRBOTR      #type complete
TITLE      trypsin (EC 3.4.21.4) precursor - bovine
CONTAINS   trypsinogen
ORGANISM   #formal_name Bos primigenius taurus #common_name cattle
DATE       24-Apr-1984 #sequence_revision 28-Feb-1986 #text_change
           18-Jul-1997
ACCESSIONS A90164; A00946; S08774
REFERENCE  A90164
#authors   Mikes, O.; Holeysovsky, V.; Tomasek, V.; Sorm, F.
```

```
#journal       Biochem. Biophys. Res. Commun. (1966) 24:346-352
#title         Covalent structure of bovine trypsinogen. The position of the
               remaining amides.
#cross-references MUID:6168848
#accession A90164
##molecule_type protein
##residues      1-57,'Q',59-67,'Q','Q',69-150,'N',152-176,'N',178-229
               ##label MIK

REFERENCE      A93755
#authors       Hartley, B.S.
#journal       Philos.Trans. R. Soc. Lond. (1970) B257:77-87
#contents      annotation: revisions
REFERENCE      A00950
#authors       Titani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
#journal       Biochemistry (1975) 14:1358-1366
#title         Amino acid sequence of dogfish trypsin.
#cross-references MUID:75146445
#contents      annotation: revisions
#note          the sequence agrees with that shown
REFERENCE      A92954
#authors       Bode, W.; Schwager, P.
#journal       J. Mol. Biol. (1975) 98:693-717
#title         The refined crystal structure of bovine beta-trypsin at 1.8
               angstrom resolution.
#cross-references MUID:76072097
#contents      annotation: X-ray crystallography: binding sites for calcium,
               substrate, and inhibitors

COMMENT        Trypsinogen is synthesized in the acinar cells of the pancreas.
COMMENT        Autocatalytic cleavage after Lys-6 leads to beta-trypsin by
               releasing a terminal hexapeptide. Subsequent cleavage after
               Lys-131 leads to alpha-trypsin. Further cleavage after Lys-176
               yields pseudotrypsin. A cleavage may also occur after Arg-105.
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS       hydrolase; pancreas; protein digestion; serine proteinase;
               zymogen
FEATURE
     1-229     #product trypsinogen #status experimental #label ZYM\
     1-6       #domain activation peptide #status experimental #label
               APT\
     7-222     #domain trypsin homology #label TRY\
     7-131,132-229 #product alpha-trypsin #status experimental #label MPT\
     6-7       #cleavage_site Lys-Ile (enteropeptidase) #status
               experimental\
     13-143,31-47,
     115-216,122-189,
     154-168,179-203  #disulfide_bonds #status experimental\
     46,90,183  #active_site His, Asp, Ser #status experimental\
     58,60,63,68 #binding_site calcium (Glu, Asn, Val, Glu) #status
               experimental\
     131-132   #cleavage_site Lys-Ser (autolytic) #status experimental
SUMMARY        #length 229 #molecular-weight 23993 #checksum 2248

Query Match     43.1%;  Score 494;  DB 1;  Length 229;
Best Local Similarity 43.1%;  Pred. No. 1.53e-86;
Matches   69;  Conservative  36;  Mismatches  52;  Indels   3;  Gaps   3;

Db     71  ISASKSIVHPSYNSNTLNNDIMLIKLKSAASLNSRVASISLFTSCASAGTQCLISGWGNT 130
           : ||:: ||:||::|||||| |||:|||||  : |: ; ||:| |:||| ||:|||
Qy      2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db    131  KSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDSCQGDSGGPVVCS 190
           :| ||| || ||: :::|: | | :||||| | :|||||:|| |||||||||:::|:
Qy     62  -ANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db    191  GKLQGIVSWGSG-CAQKNKPGVYTKVCNYVSWIKQTIASN 229
           | |||||||| | |||:| ||||:|||| ||:||
Qy    120  GYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT  14
ENTRY          T01779       #type complete
TITLE          trypsin (EC 3.4.21.4) - plaice
ORGANISM       #formal_name Pleuronectes platessa #common_name plaice
```

```
DATE           19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change
               19-Feb-1999
ACCESSIONS     T01779
REFERENCE      Z14422
#authors       Leaver, M.J.; George, S.G.
#submission    submitted to the EMBL Data Library, March 1996
#accession     T01779
##status       preliminary: translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues      1-250 ##label LEA
##cross-references EMBL:X56744; NID:g1213630; PID:g64240
KEYWORDS       hydrolase; protein digestion; serine proteinase
SUMMARY        #length 250 #molecular-weight 27705 #checksum 4482

Query Match     43.1%;  Score 494;  DB 2;  Length 250;
Best Local Similarity 44.4%;  Pred. No. 1.53e-86;
Matches   72;  Conservative  36;  Mismatches  49;  Indels   5;  Gaps   4;

Db     88  LMKIDTIIWHPNDYQTLDFDIMLIKLYHPVEVSHAVGPIPLPTSCPVAGTPCSVSGWGN 147
           ::  :::||:|: : | |||||   :|::: ||:: :|  ||: |||
Qy      1  MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL 60

Db    148  TARDGDEVYLPTLLQCMDVPIVDEEQCMKSYPDMISPRMVCAGFMDGSRDACNGDSGSPL 207
           || : ::| ::| :: || ||:||| | ; ||| |||||||: ||||||||:::
Qy     61  LA-NG-R--MPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPL 116

Db    208  VCRGEVYGLVSWGQG-CAQPNYPGVYVKLCEFLGWIERTLEA 248
           :| :| ||:| | |: | ||| ||:|:||::: ||| |::|:
Qy    117  ICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQA 158

RESULT  15
ENTRY          A35871       #type complete
TITLE          trypsin (EC 3.4.21.4) precursor, pancreatic - African clawed
               frog
ORGANISM       #formal_name Xenopus laevis #common_name African clawed frog
DATE           09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change
               08-Sep-1997
ACCESSIONS     A35871; S12117
REFERENCE      A35871
#authors       Shi, Y.B.; Brown, D.D.
#journal       Genes Dev. (1990) 4:1107-1113
#title         Developmental and thyroid hormone-dependent regulation of
               pancreatic genes in Xenopus laevis.
#cross-references MUID:91007255
#accession A35871
##status       preliminary
##molecule_type mRNA
##residues      1-243 ##label SHI
##cross-references EMBL:X53458; NID:g65162; PID:g65163
CLASSIFICATION #superfamily trypsin; trypsin homology
KEYWORDS       hydrolase; protein digestion; serine proteinase
FEATURE
     21-236    #domain trypsin homology #label TRY\
     27-157,45-61,
     129-230,136-203,  #disulfide_bonds #status predicted\
     168-182    #active_site His, Asp, Ser #status predicted\
     60,104,197 #molecular-weight 25492 #checksum 8510
SUMMARY        #length 243

Query Match     43.0%;  Score 492;  DB 2;  Length 243;
Best Local Similarity 41.9%;  Pred. No. 4.49e-86;
Matches   67;  Conservative  43;  Mismatches  47;  Indels   3;  Gaps   3;

Db     85  ISSSKVIRHSGVNSYTLDNDIMLIKLSSPASLNAAVNTVPLPSGCSAAGTSCLISGWGNT 144
           : :| ::|| ||  :|||||||||:  ::  : : : ::: :|:| |: ||:|||
Qy      2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db    145  LSNGSNYPDLLQCLNAPILTNAQCNSAYPGEITANMICVGYMEGGKDSCQGDSGGPVVCN 204
           | :||: || :|| |:| |:|  |: |||:| :|||||||: |||||||||:::|:
Qy     61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db    205  GQLQGVVSWGYG-CAMRNYPGVYTKVCNYNAWIQNTIAAN 243
```

Qy   120 GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159
          | ||:|| | : |: |||||::|:: ||::|:.|:

Search completed: Tue Sep 28 10:58:10 1999
Job time : 30 secs.

This Page Blank (uspto)

```
****************************************************************************
****************************************************************************
       _   _  ___  ___  ___  ___  _  _
      | \ | || _ \/ __|| _ \/ __|| || |
      |  \| ||  _/\__ \|   /| (__ | __ |  (TM)
      |_|\__||_|  |___/|_|_\ \___||_||_|
****************************************************************************
****************************************************************************
```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp   protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Sep 28 10:58:27 1999;   MasPar time 6.58 Seconds
                                              682.811 Million cell updates/sec

Tabular output not generated.

Title: >US-09-030-606-172
Description: (1-159) from US09030606.pep
Perfect Score: 1145
Sequence: 1 MVEASLSVRHPEYNRPLLAN.........GVYTNLCKFTEWIEKTVQAS 159

Scoring table: PAM 150
                  Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database: swiss-prot37
                  1:swissprot

Statistics: Mean 42.217;   Variance 62.196;   scale 0.679

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	543	47.4	253	1	SCCE_HUMAN	STRATUM CORNEUM CHYMOT	7.09e-114
2	519	45.3	248	1	TRY3_CHICK	TRYPSINOGEN II-P29 PRE	2.42e-107
3	516	45.1	244	1	TRY2_XENLA	TRYPSINOGEN PRECURSOR.	1.58e-106
4	516	45.1	247	1	TRY3_RAT	TRYPSINOGEN III, CATIO	1.58e-106
5	514	44.9	246	1	TRYB_RAT	TRYPSINOGEN V-B PRECUR	5.53e-106
6	512	44.7	246	1	TRYP_MOUSE	TRYPSINOGEN PRECURSOR	1.93e-105
7	506	44.2	244	1	PSS9_HUMAN	PROTEASE M PRECURSOR (	8.18e-104
8	505	44.1	246	1	TRYA_RAT	TRYPSINOGEN V-A PRECUR	1.53e-103
9	503	43.9	248	1	TRY1_CHICK	TRYPSINOGEN I-P1 PRECU	5.32e-103
10	502	43.8	248	1	TRY2_CHICK	TRYPSINOGEN I-P38 PREC	9.93e-103
11	500	43.7	247	1	TRY4_RAT	TRYPSINOGEN IV PRECURS	3.46e-102
12	497	43.4	246	1	TRY1_CANFA	TRYPSINOGEN, CATIONIC	2.24e-101
13	495	43.2	246	1	TRY1_RAT	TRYPSINOGEN I, ANIONIC	7.81e-101
14	495	43.2	247	1	TRY2_CANFA	TRYPSINOGEN, ANIONIC P	7.81e-101
15	494	43.1	243	1	TRY1_BOVIN	TRYPSINOGEN, CATIONIC	1.46e-100
16	492	43.0	243	1	TRY1_XENLA	TRYPSIN PRECURSOR (EC	5.06e-100
17	480	41.9	246	1	TRY2_RAT	TRYPSINOGEN II, ANIONI	8.83e-97
18	478	41.7	238	1	TRY3_SALSA	TRYPSINOGEN III PRECUR	3.06e-96
19	478	41.7	247	1	TRY1_HUMAN	TRYPSINOGEN I PRECURSO	3.06e-96
20	471	41.1	231	1	TRYP_PIG	TRYPSIN PRECURSOR (EC	2.36e-94
21	470	41.0	261	1	KLKA_MOUSE	GLANDULAR KALLIKREIN K	4.38e-94
22	468	40.9	247	1	TRY2_BOVIN	TRYPSINOGEN, ANIONIC P	1.51e-93
23	462	40.3	261	1	KLKC_MOUSE	GLANDULAR KALLIKREIN K	6.23e-92
24	460	40.2	229	1	TRYP_SQUAC	TRYPSIN PRECURSOR (EC	2.15e-91
25	460	40.2	242	1	TRY1_SALSA	TRYPSINOGEN I PRECURS	2.15e-91
26	456	39.8	231	1	TRY2_SALSA	TRYPSINOGEN II PRECURS	2.55e-90
27	454	39.7	261	1	KLKZ_MOUSE	GLANDULAR KALLIKREIN K	8.79e-90
28	455	39.7	304	1	TRY4_HUMAN	TRYPSINOGEN IVA PRECUR	4.74e-90
29	453	39.6	188	1	KLK_RAT	GLANDULAR KALLIKREIN 3	1.63e-89
30	449	39.2	241	1	TRYX_GADMO	TRYPSINOGEN X PRECURSO	1.93e-88
31	449	39.2	261	1	KLK3_MOUSE	GLANDULAR KALLIKREIN K	1.93e-88
32	446	39.0	261	1	KLK9_MOUSE	GLANDULAR KALLIKREIN K	1.23e-87
33	445	38.9	244	1	KLKA_RAT	GLANDULAR KALLIKREIN 1	2.28e-87
34	444	38.8	247	1	TRY2_HUMAN	TRYPSINOGEN II PRECURS	4.22e-87
35	440	38.4	261	1	KLK8_RAT	GLANDULAR KALLIKREIN 8	4.97e-86
36	439	38.3	259	1	KLKL_MOUSE	GLANDULAR KALLIKREIN 1	9.20e-86
37	438	38.3	259	1	KLKB_RAT	GLANDULAR KALLIKREIN 1	1.70e-85
38	437	38.2	241	1	TRY1_GADMO	TRYPSINOGEN I PRECURSO	3.15e-85
39	434	37.9	261	1	KLK6_MOUSE	GLANDULAR KALLIKREIN K	2.00e-84
40	427	37.3	261	1	KLK5_MOUSE	GLANDULAR KALLIKREIN K	1.48e-82
41	426	37.2	261	1	KLK7_RAT	GLANDULAR KALLIKREIN 7	2.73e-82
42	425	37.1	261	1	KLK1_RAT	GLANDULAR KALLIKREIN,	5.04e-82
43	422	36.9	261	1	KLK1_MOUSE	GLANDULAR KALLIKREIN K	3.18e-81
44	420	36.7	247	1	TRY3_HUMAN	TRYPSINOGEN III PRECUR	1.08e-80
45	418	36.5	259	1	KLK2_RAT	TONIN PRECURSOR (EC 3.	3.69e-80

ALIGMENTS

RESULT   1
ID   SCCE_HUMAN   STANDARD;   PRT;   253 AA.
AC   P49862;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE).
GN   PRSS6 OR SCCE.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC   TISSUE=SKIN;
RX   MEDLINE; 94308225.
RA   HANSSON L., STROEMQVIST M., BAECKMAN A., WALLBRANDT P., CARLSTEIN A.,
RA   EGELRUD T.;
RT   "Cloning, expression, and characterization of stratum corneum
RT   chymotryptic enzyme. A skin-specific human serine proteinase.";
RL   J. BIOL. CHEM. 269:19420-19426(1994).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE; 95314630.
RA   SKYTT A., STROEMQVIST M., EGELRUD T.;
RT   "Primary substrate specificity of recombinant human stratum corneum
RT   chymotryptic enzyme.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 211:586-589(1995).
-!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
CC   STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
CC   SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
CC   RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CC   CLEAVES INSULIN B CHAIN AT 6-LEU-|-CYS-7, 16-TYR-|-LEU-17, 25-PHE-
CC |-TYR-26, AND 26-TYR-|-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
CC   OF PRECURSORS TO INFLAMMATORY CYTOKINES.
-!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
CC   EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
CC   ALSO SEEN IN THE BRAIN AND KIDNEY.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC   TRYPSIN FAMILY.

CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

```
CC  ------
DR  EMBL; L33404; G532504; -.
DR  PROSITE; PS00134; TRYPSIN_HIS; 1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
DR  PFAM; PF00089; trypsin; 1.
DR  HSSP; P00763; 1DPO.
KW  HYDROLASE; SERINE PROTEASE; ZYMOGEN; GLYCOPROTEIN; SIGNAL.
FT  SIGNAL         1     22
FT  PROPEP        23     29       ACTIVATION PEPTIDE.
FT  CHAIN         30    253       STRATUM CORNEUM CHYMOTRYPTIC ENZYME.
FT  ACT_SITE      70     70       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE     112    112       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE     205    205       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  DISULFID      36    137       BY SIMILARITY.
FT  DISULFID      55     71       BY SIMILARITY.
FT  DISULFID     144    211       BY SIMILARITY.
FT  DISULFID     176    190       BY SIMILARITY.
FT  DISULFID     201    226       BY SIMILARITY.
FT  CARBOHYD     246    246       POTENTIAL.
SQ  SEQUENCE     253 AA;  27525 MW;  07FDB9F7 CRC32;

Query Match             47.4%;
Best Local Similarity   46.5%;  Score 543;  DB 1; Length 253;
Matches   73;  Conservative  29;  Mismatches 53;  Indels   2;  Gaps   2;

Db   93  IKASKSFRHPGYSTQTHVNDLMIVKLNSQARLSSMVKKVRLPSRCEPPGTTCTVSGWGTT 152
             : |:|| | ||:   | |||||||:  :  : :|:||   :| :|:| | |||||
Qy    2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db  153  TSPDVTFPSDLMCVDVKLISPQDCTKVYKDLLENSMLCAGIPDSKKNACNGDSGGPLVCR 212
            :  | :|||:  ||   ::|: |:|:|    |:|||  :||||||||||||||||||||
Qy   62  AN-G-RMPIVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db  213  GILQGLVSWGTFPCGQPNDPGVYTQVCKFTKWINDTM 249
           | ||||| | ||| :| |||| |||||:|
Qy  120  GILQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTV 156


RESULT  2
ID  TRY3_CHICK   STANDARD;      PRT;      248 AA.
AC  Q90629;
DT  01-NOV-1997 (REL. 35, CREATED)
DT  01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE  TRYPSINOGEN II-P29 PRECURSOR.
OS  GALLUS GALLUS (CHICKEN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC  NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=PANCREAS;
RX  MEDLINE; 95251611.
RA  WANG K., GAN L., LEE I., HOOD L.E.;
RT  "Isolation and characterization of the chicken trypsinogen gene
RT  family.";
RL  BIOCHEM. J. 307:471-479(1995).
CC  -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC  -!- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
CC      LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPSIN FAMILY.
```

```
DR  EMBL; U15157; G603907; -.
DR  PROSITE; PS00134; TRYPSIN_HIS; 1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
DR  PFAM; PF00089; trypsin; 1.
DR  HSSP; P00763; 1DPO.
KW  HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW  MULTIGENE FAMILY.
FT  SIGNAL         1     16       BY SIMILARITY.
FT  PROPEP        17     25       ACTIVATION PEPTIDE (BY SIMILARITY).
FT  CHAIN         26    248       TRYPSIN II-P29.
FT  ACT_SITE      65     65       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE     109    109       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  ACT_SITE     202    202       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT  DISULFID      32    162       BY SIMILARITY.
FT  DISULFID      50     66       BY SIMILARITY.
FT  DISULFID     134    235       BY SIMILARITY.
FT  DISULFID     141    208       BY SIMILARITY.
FT  DISULFID     173    187       BY SIMILARITY.
FT  DISULFID     198    222       BY SIMILARITY.
FT  SITE         196    196       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ  SEQUENCE     248 AA;  26622 MW;  F7FE97E9 CRC32;

Query Match             45.3%;
Best Local Similarity   45.2%;  Score 519;  DB 1; Length 248;
Matches   71;  Conservative  37;  Mismatches 46;  Indels   3;  Gaps   3;

Db   92  SSVIIRHPKYSSTLNNDIMLIKLASAVEYSADIQPIALPSSCAKAGTECLLISGWGNTLS 151
            :: :||| : | |||||||  :| | |:::::|:|::  |||:|:|||||||
Qy    4  ASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-LLA 62

Db  152  NGYNYPELLQCLNAPILSDQECOEAYPGDITSNMICVGFLEGGKDSCQGDSGGPVVCNGE 211
          || | ||||||   ::| :  :::| : ||||||:::|||
Qy   63  NG-RMPIVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGY 121

Db  212  LQGIVSWGIG-CALKGYPGVYTKVCNVVDWIQETIAA 247
          ||||| | :| | ||||||||||| ::|: :|:
Qy  122  LQGIVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQA 158


RESULT  3
ID  TRY2_XENLA   STANDARD;      PRT;      244 AA.
AC  P70059;
DT  01-NOV-1997 (REL. 35, CREATED)
DT  01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT  01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE  TRYPSINOGEN PRECURSOR.
OS  XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC  MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  WANG K., LYTLE L., GAN L., HOOD L.E.;
RL  SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC  -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC  -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPSIN FAMILY.
```

```
DR  EMBL; U72330; G1621633; -.
DR  PROSITE; PS00134; TRYPSIN_HIS; 1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
DR  PFAM; PF00089; trypsin; 1.
DR  HSSP; P00763; 1DPO.
KW  HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW  MULTIGENE FAMILY.
FT  SIGNAL         1     15       BY SIMILARITY.
FT  PROPEP        16     21       ACTIVATION PEPTIDE (BY SIMILARITY).
```

```
FT   CHAIN        22    244       TRYPSIN.
FT   ACT_SITE     61     61       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    105    105       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    198    198       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     28    158       BY SIMILARITY.
FT   DISULFID     46     62       BY SIMILARITY.
FT   DISULFID    137    204       BY SIMILARITY.
FT   DISULFID    169    183       BY SIMILARITY.
FT   DISULFID    194    218       BY SIMILARITY.
FT   SITE        192    192       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE    244 AA;   26079 MW;  754B2E0E CRC32;

Query Match           45.1%;   Score 516;  DB 1;  Length 244;
Best Local Similarity 43.8%;   Pred. No. 1.58e-106;
Matches   70;  Conservative 40;  Mismatches 47;  Indels  3;  Gaps  3;

Db   86  IDSQKVIKHPNYNSRNLDNDIMLIKLSTTARLSANIQSVPLPSACASAGTNCLLISGWGNT 145
             ::  :||: ||   ||||||| ::::|||:|:::||||
Qy    2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db  146  LSSGTNYPDLLQCLNAPILTDSQCNSYPGEITKNMFCAGFLAGGKDSCQGDSGGPVVCN 205
            |:: |:|||:| :::: ||:|        ||| ||||||:|
Qy   61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db  206  GQLQGVSWGYG-CAQRYNPGVYTKVCNFVTWIQSTISSN 244
            ||||:|| | :|| ||||||:::|:| || :::
Qy  120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT 4
ID   TRY3_RAT     STANDARD;    PRT;   247 AA.
AC   P08426;
DT   01-AUG-1988 (REL. 08, CREATED)
DT   01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN III, CATIONIC PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN
DE   III).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87271609.
RA   FLETCHER T.S., ALHADEFF M., CRAIK C.S., LARGMAN C.;
RT   "Isolation and characterization of a cDNA encoding rat cationic
RT   trypsinogen.";
RL   BIOCHEMISTRY 26:3081-3086(1987).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; M16624; G206499; -.
DR   PIR; A27547; A27547.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15
FT   PROPEP       16     24       ACTIVATION PEPTIDE.
FT   CHAIN        25    247       TRYPSIN III, CATIONIC.
FT   ACT_SITE     64     64       CHARGE RELAY SYSTEM (BY SIMILARITY).
```

```
FT   ACT_SITE    108    108       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    201    201       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     31     49       BY SIMILARITY.
FT   DISULFID    133    234       BY SIMILARITY.
FT   DISULFID    140    207       BY SIMILARITY.
FT   DISULFID    172    186       BY SIMILARITY.
FT   DISULFID    197    221       BY SIMILARITY.
FT   SITE        195    195       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE    247 AA;   26269 MW;  7DD35F0B CRC32;

Query Match           45.1%;   Score 516;  DB 1;  Length 247;
Best Local Similarity 43.1%;   Pred. No. 1.58e-106;
Matches   69;  Conservative 41;  Mismatches 47;  Indels  3;  Gaps  3;

Db   89  IDAAKIIRHPSYNANTFDNDIMLIKLNSPATLNSRVSTVSLPRSCGSSGTKCLVSGWGNT 148
             :::  :|| |    ||||||| ::::|||
Qy    2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db  149  LSSGTNYPSLLQCLDAPVLSDSSCKSSYPGKITSNMFCLGFLEGGKDSCQGDSGGPVVCN 208
            |:: |:|||:|: :|: : ||:|        ||| ||||||:|
Qy   61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db  209  GQLQGVSWGYG-CAQKGKPGVYTKVCNYVNWIQQTVAAN 247
            ||||:|| | :|| ||||||:::|: || :::
Qy  120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT 5
ID   TRYB_RAT     STANDARD;    PRT;   246 AA.
AC   P32822;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN V-B PRECURSOR (EC 3.4.21.4).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 92165057.
RA   KANG J., WIEGAND U., MUELLER-HILL B.;
RT   "Identification of cDNAs encoding two novel rat pancreatic serine
RT   proteases.";
RL   GENE 110:181-187(1992).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; X59013; G57415; -.
DR   PIR; JQ1472; JQ1472.
DR   PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15
FT   PROPEP       16     24       ACTIVATION PEPTIDE.
FT   CHAIN        25    246       TRYPSIN V-B.
FT   ACT_SITE     64     64       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    108    108       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    200    200       CHARGE RELAY SYSTEM (BY SIMILARITY).
```

FT   DISULFID   31   160   BY SIMILARITY.
FT   DISULFID   49   65   BY SIMILARITY.
FT   DISULFID   133   233   BY SIMILARITY.
FT   DISULFID   140   206   BY SIMILARITY.
FT   DISULFID   171   185   BY SIMILARITY.
FT   DISULFID   196   220   BY SIMILARITY.
FT   SITE   194   194   REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE   246 AA; 26819 MW; 743CC84B CRC32;

Query Match   44.9%; Score 514; DB 1; Length 246;
  Best Local Similarity   44.0%; Pred. No. 5.53e-106;
  Matches   70; Conservative   39; Mismatches   48; Indels   2; Gaps   2;

Db   89 IDAAKMILHPDYDKWTVDNDIMLIKLKSPATLNSKVSTIPLPQYCPTAGTECLVSGWGVL 148
              ::|: :|:|::: |:|||| ::||||| |||||:|
Qy   2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL 61

Db   149 KFGFESPSVLQCLDAPVLSDSVCHKAYPRQITNNMFCLGFLEGGKDSCQYDSGGPVVCNG 208
              | ||||:||: :|::: | |: :|| ||||:|||
Qy   62 ANG-RMPTVLQCVNVSVSSEEVCSKLYDPLYHPSMFCAGGGQXXDSCNGDSGGPLICNG 120

Db   209 EVQGIVSWGDG-CALEGKPGVTKVCNYLNWIQQTVAAN 246
              |:|||:|| | || |||||:|||| ||: :|| | |
Qy   121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 6
ID   TRYP_MOUSE   STANDARD;   PRT;   246 AA.
AC   P07146;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN PRECURSOR (EC 3.4.21.4).
GN   TRY2.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/J;
RX   MEDLINE; 87066713.
RA   STEVENSON B.J., HAGENBUCHLE O., WELLAUER P.K.;
RT   "Sequence organisation and transcriptional regulation of the mouse
RT   elastase II and trypsin genes.";
RL   NUCLEIC ACIDS RES. 14:8307-8330(1986).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC   TRYPSIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X04574; G54919; -.
DR   EMBL; X04577; G54917; -.
DR   PIR; B25528; B25528.
DR   MGD; MGI:102759; TRY2.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL.
FT   SIGNAL   1   15
FT   PROPEP   16   23   ACTIVATION PEPTIDE.
FT   CHAIN   24   246   TRYPSIN.
FT   ACT_SITE   63   63   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   107   107   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   200   200   CHARGE RELAY SYSTEM (BY SIMILARITY).

FT   DISULFID   30   160   BY SIMILARITY.
FT   DISULFID   48   64   BY SIMILARITY.
FT   DISULFID   132   233   BY SIMILARITY.
FT   DISULFID   139   206   BY SIMILARITY.
FT   DISULFID   171   185   BY SIMILARITY.
FT   DISULFID   196   220   BY SIMILARITY.
FT   SITE   194   194   REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE   246 AA; 26203 MW; BD975983 CRC32;

Query Match   44.7%; Score 512; DB 1; Length 246;
  Best Local Similarity   43.9%; Pred. No. 1.93e-105;
  Matches   69; Conservative   39; Mismatches   46; Indels   3; Gaps   3;

Db   88 VDSAKIIRHPNYNSWTLDNDIMLIKLASPVTLNARVASVPLPSSCAPAGTQCLISGWGNT 147
              :::: ::|||:|| | ||:|||| ::|: :|||||| |||||||
Qy   2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db   148 LSNGVNNPDLIQCVDAPVLPQADCEASYPGDITNNMICVGFLEGGKDSCQGDSGGPVVCN 207
              :|:| |:|||::|:::: | |:|:|| ||||:||:
Qy   61 LANG-RMPTVLQCVNVSVSSEEVCSKLYDPLYHPSMFCAGGGQXXDSCNGDSGGPLICN 119

Db   208 GELQGIVSWGYG-CAQPDAPGVTKVCNYVDWIQNTI 243
              | |||:|| | | |||||:||| | ::||
Qy   120 GYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTV 156

RESULT 7
ID   PSS9_HUMAN   STANDARD;   PRT;   244 AA.
AC   Q9876;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROTEASE M PRECURSOR (EC 3.4.21.-) (NEUROSIN) (ZYME) (SP59).
GN   PRSS9.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97053999.
RA   ANISOWICZ A., SOTIROPOULOU G., STENMAN G., MOK S.C., SAGER R.;
RT   "A novel protease homolog differentially expressed in breast and
RT   ovarian cancer.";
RL   MOL. MED. 2:624-636(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RX   MEDLINE; 97157069.
RA   YAMASHIRO K., TSURUOKA N., KODAMA S., TSUJIMOTO M., YAMAMURA Y.,
RA   TANAKA T., NAKAZATO H., YAMAGUCHI N.;
RT   "Molecular cloning of a novel trypsin-like serine protease (neurosin)
RT   preferentially expressed in brain.";
RL   BIOCHIM. BIOPHYS. ACTA 1350:11-14(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE; 97460104.
RA   LITTLE S.P., DIXON E.P., NORRIS F., BUCKLEY W., BECKER G.W.,
RA   JOHNSON M., DOBBINS J.R., WYRICK T., MILLER J.R., MACKELLAR W.,
RA   HEPBURN D., CORVALAN J., MCCLURE D., LIU X., STEPHENSON D.,
RA   CLEMENS J., JOHNSTONE E.M.;
RT   "Zyme, a novel and potentially amyloidogenic enzyme cDNA isolated
RT   from Alzheimer's disease brain.";
RL   J. BIOL. CHEM. 272:25135-25142(1997).
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN BRAIN. ALSO FOUND
CC   IN COLON AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC   TRYPSIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its

CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U62801; G1518788; -.
DR   EMBL; D78203; G1805493; -.
DR   EMBL; AF013988; G2318115; -.
DR   MIM; 602652; -.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
KW   HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ZYMOGEN; SIGNAL.
FT   SIGNAL   1   16   POTENTIAL.
FT   PROPEP   17   21   ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN   22   244   PROTEASE M.
FT   ACT_SITE   62   62   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   106   106   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   197   197   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID   28   157   BY SIMILARITY.
FT   DISULFID   47   63   BY SIMILARITY.
FT   DISULFID   131   231   BY SIMILARITY.
FT   DISULFID   138   203   BY SIMILARITY.
FT   DISULFID   168   182   BY SIMILARITY.
FT   DISULFID   193   218   BY SIMILARITY.
FT   CARBOHYD   134   134   POTENTIAL.
SQ   SEQUENCE   244 AA;   26856 MW;   0CFCEFCE CRC32;

Query Match   44.2%;
Best Local Similarity   42.1%;   Score 506; DB 1; Length 244;
Matches   64; Conservative 33; Mismatches 55; Indels 0; Gaps 0;

Db   92 AVIHPDYDAASHDQDIMLLRLARPAKLSELIQPLPERDCSANTTSCHILGWGKTADGDF 151
          ::|::|| :::| | |: :: : |::::| ||| : ||| | ::|: || |||
Qy   7 SVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWLLANGRM   66

Db   152 PDTIOCAYIHLVSRECEHAYPGQITQNMLCAGDEKYGKDSCQGDSGGPLVCGDHLRGLV 211
          : |:| ||| :|| | : || ||:|||||||||||::| |:|||
Qy   67 PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICNGYLQGLV 126

Db   212 SWGNIPCGSKEKPGVYTNVCRYTNWIQKTIQA 243
          :|:||| |||||||||::|:|:|:|
Qy   127 SFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQA 158

RESULT 8
ID   TRYA_RAT   STANDARD;   PRT;   246 AA.
AC   P32821;
DT   01-OCT-1993 (REL. 27, CREATED)
DT   01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN V-A PRECURSOR (EC 3.4.21.4).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 92165057.
RA   KANG J., WIEGAND U., MUELLER-HILL B.;
RT   "Identification of cDNAs encoding two novel rat pancreatic serine
RT   proteases.";
RL   GENE 110:181-187(1992).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC   TRYPSIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

---

CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; X59012; G57413; -.
DR   PIR; JQ1471; JQ1471.
DR   PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL   1   15   ACTIVATION PEPTIDE.
FT   PROPEP   16   24   TRYPSIN V-A.
FT   CHAIN   25   246   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   64   64   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   108   108   CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   200   200   BY SIMILARITY.
FT   DISULFID   31   160   BY SIMILARITY.
FT   DISULFID   49   65   BY SIMILARITY.
FT   DISULFID   133   233   BY SIMILARITY.
FT   DISULFID   140   206   BY SIMILARITY.
FT   DISULFID   171   185   BY SIMILARITY.
FT   DISULFID   196   220   BY SIMILARITY.
FT   SITE   194   194   REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE   246 AA;   26900 MW;   EA6C939B CRC32;

Query Match   44.1%;
Best Local Similarity   43.6%;   Score 505; DB 1; Length 246;
Matches   68; Conservative 38; Mismatches 48; Indels 2; Gaps 2;

Db   89 IDAAKMLHPDYDKWTVDNDIMLIKLKSPATLNSKVSTIPLPQYCPPAGTECLVSGWGVL 148
          ::|: :||:||| |:|:|||| |:||||||: : |::: :   :| |||||
Qy   2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWLL   61

Db   149 KFGFESPSVLQCLDAPVLSDSVCHKAYPRQITNNMFCLGFLEGGKDSCQYDSGGPVVCNG 208
          | :|||||: :|:|||| : |:::: ||| |||::::| |||::|||
Qy   62 ANG-RMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGQXQXDSCNGDSGGPLICNG 120

Db   209 EVQGIVSWGDG-CALEGKPGVYTKVCNYLNWIHQTI 243
          :|| :|||| | :|::||||||:| | |:||:: |:
Qy   121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTV 156

RESULT 9
ID   TRY1_CHICK   STANDARD;   PRT;   248 AA.
AC   Q90627;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN I-P1 PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 95251611.
RA   WANG K., GAN L., LEE I., HOOD L.E.;
RT   "Isolation and characterization of the chicken trypsinogen gene
RT   family.";
RL   BIOCHEM. J. 307:479-479(1995).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
CC   LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC   TRYPSIN FAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U15155; G603903; -.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
     HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15       BY SIMILARITY.
FT   PROPEP       16     25       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN        26    248       TRYPSIN I-P1.
FT   ACT_SITE     65     65       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    109    109       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    202    202       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     32    162       BY SIMILARITY.
FT   DISULFID     50     66       BY SIMILARITY.
FT   DISULFID    134    235       BY SIMILARITY.
FT   DISULFID    141    208       BY SIMILARITY.
FT   DISULFID    173    187       BY SIMILARITY.
FT   DISULFID    198    222       BY SIMILARITY.
FT   SITE        196    196       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE  248 AA;  26069 MW;  98D292B2 CRC32;

Query Match             43.9%;  Score 503;  DB 1;  Length 248;
Best Local Similarity   43.1%;  Pred. No. 5.32e-103;  Indels 3;  Gaps 3;
Matches   69;  Conservative   38;  Mismatches   50;

Db    90  ISSSKVIRHSGYNANTLNNDIMLIKLSKAATLNSYVNTVPLPTSCVTAGTTCLISGWGNT 149
              :: :||: ||  |  |||||| | ||:|||  : :::::| |||  :|||:|||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L  60

Db   150  LSSGSLYPDVLQCLNAPVLSSSQCSSAYPGRITSNMICIGYLNGGKDSCQGDSGGPVVCN 209
              |::| || |||:  ||:|: :::||   ::|||||||:  :|||::|
Qy    61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLIHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   210  GQLQGIVSWGIG-CAQKGYPGVYTKVCNYVSWIKTTMSSN 248
              | |||:|||| : :| | | |||:|::| | :: ::
Qy   120  GYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 10
ID   TRY2_CHICK     STANDARD;      PRT;   248 AA.
AC   Q90628;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN I-P38 PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RX   MEDLINE; 95251611.
RA   WANG K., GAN L., LEE I., HOOD L.E.;
RT   "Isolation and characterization of the chicken trypsinogen gene
RT   family.";
RL   BIOCHEM. J. 307:471-479(1995).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE
CC       LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U15156; G603905; -.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
     HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15       BY SIMILARITY.
FT   PROPEP       16     25       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN        26    248       TRYPSIN I-P38.
FT   ACT_SITE     65     65       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    109    109       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    202    202       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     32    162       BY SIMILARITY.
FT   DISULFID     50     66       BY SIMILARITY.
FT   DISULFID    134    235       BY SIMILARITY.
FT   DISULFID    141    208       BY SIMILARITY.
FT   DISULFID    173    187       BY SIMILARITY.
FT   DISULFID    198    222       BY SIMILARITY.
FT   SITE        196    196       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE  248 AA;  26087 MW;  F64E0643 CRC32;

Query Match             43.8%;  Score 502;  DB 1;  Length 248;
Best Local Similarity   43.1%;  Pred. No. 9.93e-103;  Indels 3;  Gaps 3;
Matches   69;  Conservative   38;  Mismatches   50;

Db    90  ISSSKVIRHSGYNANTLNNDIMLIKLSKAATLNSYVNTVPLPTSCVTAGTTCLISGWGNT 149
              :: :||: ||  |  |||||| | ||:|||  : :::::| |||  :|||:|||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L  60

Db   150  LSSGSLYPDVLQCLNAPVLSSSQCSSAYPGRITSNMICIGYLNGGKDSCQGDSGGPVVCN 209
              |::| || |||:  ||:|: :::||   ::|||||||:  :|||::|
Qy    61  LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLIHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   210  GQLQGFVSWGIG-CAQKGYPGVYTKVCNYVSWIKTTMSSN 248
              | |||:|||| : :| | | |||:|::| | :: ::
Qy   120  GYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 11
ID   TRY4_RAT       STANDARD;      PRT;   247 AA.
AC   P12788;
DT   01-OCT-1989 (REL. 12, CREATED)
DT   01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN IV PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN IV).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=PANCREAS;
RX   MEDLINE; 89386010.
RA   LUETCKE H.A., RAUSCH U., VASILOUDES P., SCHEELE G.A., KERN H.F.;
RT   "A fourth trypsinogen (P23) in the rat pancreas induced by CCK.";
RL   NUCLEIC ACIDS RES. 17:6736-6736(1989).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- INDUCTION: BY CCK.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X15679; G56814; -.
```

```
DR   PIR; S05494; S05494.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15
FT   PROPEP       16     23                ACTIVATION PEPTIDE.
FT   CHAIN        24    247                TRYPSIN IV.
FT   ACT_SITE     64     64                CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    108    108                CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    201    201                CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     30    161                BY SIMILARITY.
FT   DISULFID     49     65                BY SIMILARITY.
FT   DISULFID    133    234                BY SIMILARITY.
FT   DISULFID    140    207                BY SIMILARITY.
FT   DISULFID    172    186                BY SIMILARITY.
FT   DISULFID    197    221                BY SIMILARITY.
FT   SITE        195    195                REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE    247 AA;   26573 MW;   258664D7 CRC32;

Query Match             43.7%;     Score 500;   DB 1;   Length 247;
Best Local Similarity   42.0%;     Pred. No. 3.46e-102;
Matches       66;   Conservative   39;   Mismatches   49;   Indels    3;   Gaps    3;

Db    89  IDAEKIIRHPEYNKDTLDNDIMLIKLKSPAVLNSQVSTVSLPRSCASTDAQCLVSGWGNT  148
              :|| ||||||:||  ||||||||||| :|||||||| :||::||:   |:|||||||||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L   60

Db   149  VSIGGKYPALLQCLEAPVLSASSCKKSYPGQITSNMFCLGFLEGGKDSCDGDSGGPVVCN  208
              :|: |||| |:|  :| |     :||   | ||| ||| :||||||:||:|||||| ||
Qy    61  LA-NGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN  119

Db   209  GEIQGIVSWGSV-CAMRGKPGVYTKVCNYLSWIQETM  244
              ::|||||| |  :| |||||||| :|  |||:||:|
Qy   120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTV  156


RESULT  12
ID   TRY1_CANFA    STANDARD;      PRT;      246 AA.
AC   P06871;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN, CATIONIC PRECURSOR (EC 3.4.21.4).
OS   CANIS FAMILIARIS (DOG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86284628.
RA   PINSKY S.D., LAFORGE K.S., SCHEELE G.;
RT   "Differential regulation of trypsinogen mRNA translation: full-length
RT   mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT   in the dog pancreas.";
RL   MOL. CELL. BIOL. 5:2669-2676(1985).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M11590; G164097; -.
DR   PIR; B26273; TRDGC.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
```

```
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00761; 1EPT.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15
FT   PROPEP       16     23                ACTIVATION PEPTIDE.
FT   CHAIN        24    246                TRYPSIN, CATIONIC.
FT   ACT_SITE     63     63                CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    107    107                CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    200    200                CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     30    160                BY SIMILARITY.
FT   DISULFID     48     64                BY SIMILARITY.
FT   DISULFID    132    233                BY SIMILARITY.
FT   DISULFID    139    206                BY SIMILARITY.
FT   DISULFID    171    185                BY SIMILARITY.
FT   DISULFID    196    220                BY SIMILARITY.
FT   SITE        194    194                REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE    246 AA;   26170 MW;   DD106F5D CRC32;

Query Match             43.4%;     Score 497;   DB 1;   Length 246;
Best Local Similarity   42.5%;     Pred. No. 2.24e-101;
Matches       68;   Conservative   38;   Mismatches   51;   Indels    3;   Gaps    3;

Db    88  INAAKIIRHPRYNANTIDNDIMLIKLSSPATLNSRVSAIALPKSCPAAGTQCLISGWGNT  147
              ::|: :||| |::|:|:|||||||||||||:||:|:||:|||:   |:||  ||||||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L   60

Db   148  QSIGQNYPDVLQCLKAPILSDSVCRNAYPGQISSNMWCLGYMEGGKDSCQGDSGGPVVCN  207
              :|| ::| |||||:|||| :: :: ::|  | ||| ::||||||||||:: ::|| ||
Qy    61  LANGRM-PTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN  119

Db   208  GELQGIVSWG-AGCAQKGKPGVSPKVCKYVSWIQQTIAAN  246
              ::| ||| |  :|| ||||| |:  ||||:||::| |:
Qy   120  GYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS  159


RESULT  13
ID   TRY1_RAT      STANDARD;      PRT;      246 AA.
AC   P00762;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN I, ANIONIC PRECURSOR (EC 3.4.21.4) (PRETRYPSINOGEN I).
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 82265624.
RA   MCDONALD R.J., STARY S.J., SWIFT G.H.;
RT   "Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide
RT   sequences of the cloned cDNAs.";
RL   J. BIOL. CHEM. 257:9724-9732(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85054880.
RA   CRAIK C.S., CHOO Q.L., SWIFT G.H., QUINTO C., MCDONALD R.J.,
RA   RUTTER W.J.;
RT   "Structure of two related rat pancreatic trypsin genes.";
RL   J. BIOL. CHEM. 259:14255-14264(1984).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX   MEDLINE; 87292123.
RA   SPRANG S., STANDING T., FLETTERICK R.J., STROUD R.M., FINER-MOORE J.,
RA   XUONG N.-H., HAMLIN R., RUTTER W.J., CRAIK C.S.;
RT   "The three-dimensional structure of Asn102 mutant of trypsin: role of
RT   Asp102 in serine protease catalysis.";
RL   SCIENCE 237:905-909(1987).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- THIS SEQUENCE REPRESENTS THE PRECURSOR OF THE MAJOR FORM OF
```

```
CC       TRYPSIN PRODUCED BY THE ADULT PANCREAS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; V01273; G57408; -.
DR   EMBL; J00778; G206508; -.
DR   PIR; A00948; TRRT1.
DR   PDB; 1TRM; 15-JUL-93.
DR   PDB; 2TRM; 16-JUL-88.
DR   PDB; 1BRA; 30-APR-94.
DR   PDB; 1BRB; 31-JUL-94.
DR   PDB; 1BRC; 31-MAY-94.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY; 3D-STRUCTURE.
FT   SIGNAL        1     15
FT   PROPEP       16     23                 ACTIVATION PEPTIDE.
FT   CHAIN        24    246                 TRYPSIN I, ANIONIC.
FT   ACT_SITE     63     63                 CHARGE RELAY SYSTEM.
FT   ACT_SITE    107    107                 CHARGE RELAY SYSTEM.
FT   ACT_SITE    200    200                 CHARGE RELAY SYSTEM.
FT   DISULFID     30    160
FT   DISULFID     48     64
FT   DISULFID    132    233
FT   DISULFID    139    206
FT   DISULFID    171    185
FT   DISULFID    196    220
FT   SITE        194    194                 REQUIRED FOR SPECIFICITY.
FT   STRAND       25     25
FT   STRAND       28     29
FT   TURN         32     33
FT   TURN         36     37
FT   STRAND       38     42
FT   STRAND       46     54
FT   TURN         55     56
FT   STRAND       57     60
FT   HELIX        62     64
FT   STRAND       70     73
FT   STRAND       77     77
FT   TURN         78     79
FT   STRAND       86     95
FT   TURN         97     98
FT   TURN        100    100
FT   TURN        101    104
FT   STRAND      105    105
FT   TURN        106    106
FT   STRAND      109    113
FT   TURN        127    127
FT   TURN        135    136
FT   STRAND      138    143
FT   STRAND      157    157
FT   HELIX       159    165
FT   HELIX       168    174
FT   TURN        176    178
FT   TURN        181    182
FT   STRAND      183    186
FT   TURN        189    190
FT   STRAND      194    194
FT   TURN        197    198
FT   TURN        200    201
FT   STRAND      203    206
FT   TURN        207    208
FT   STRAND-     209    216
```

```
FT   TURN        223    224
FT   STRAND      227    231
FT   HELIX       232    234
FT   HELIX       236    244
FT   TURN        245    245
SQ   SEQUENCE   246 AA;  25959 MW;  C6D097E0 CRC32;

Query Match                    43.2%;  DB 1; Length 246;
Best Local Similarity 43.1%;   Pred. No. 7.81e-101;
Matches    69;  Conservative  38; Mismatches 50; Indels  3; Gaps  3;

Db   88 INAAKIIKHPNVSSWTLNNDIMLIKLSSPVKLNARVAPVALPSACAPAGTQCLISGWGNT 147
         ::|: .::||:|: . | ||:|||| .:| .| :::|:|| ||.:::||:|||
Qy    2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db  148 LSNGVNNPDLLQCVDAPVLSQADCEAAYPGEITSSMICVGFLEGGKDSCQGDSGGPVVCN 207
         :|:| .::|||| || | :||:| || | .||.| ::| | ::| | | ::||:||
Qy   61 LANG-RMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db  208 GQLQGIVSWGYG-CALPDNPGVYTKVCNFVGWIQDTIAAN 246
         |.||| :|:|| : |||||:.||| ||:|:|:
Qy  120 GYLQGLVSFGKAPCGQVGPVGVYTNLCKFTEWIEKTVQAS 159

RESULT   14
ID   TRY2_CANFA    STANDARD;    PRT;   247 AA.
AC   P06872;
DT   01-JAN-1988 (REL. 06, CREATED)
DT   01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN, ANIONIC PRECURSOR (EC 3.4.21.4).
OS   CANIS FAMILIARIS (DOG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86284628.
RA   PINSKY S.D., LAFORGE K.S., SCHEELE G.;
RT   "Differential regulation of trypsinogen mRNA translation: full-length
RT   mRNA sequences encoding two oppositely charged trypsinogen isoenzymes
RT   in the dog pancreas.";
RL   MOL. CELL. BIOL. 5:2669-2676(1985).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M11589; G164095; -.
DR   PIR; A26273; TRDG.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
DR   HSSP; P00763; 1DPO.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   MULTIGENE FAMILY.
FT   SIGNAL        1     15
FT   PROPEP       16     23                 ACTIVATION PEPTIDE.
FT   CHAIN        24    247                 TRYPSIN, ANIONIC.
FT   ACT_SITE     63     63                 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    107    107                 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    200    200                 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DISULFID     30    160                 BY SIMILARITY.
FT   DISULFID     48     64                 BY SIMILARITY.
FT   DISULFID    132    233                 BY SIMILARITY.
FT   DISULFID    139    206                 BY SIMILARITY.
```

```
FT   DISULFID    171    185         BY SIMILARITY.
FT   DISULFID    196    220         BY SIMILARITY.
FT   SITE        194    194         REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ   SEQUENCE    247 AA;  26423 MW;  59BC6877 CRC32;

Query Match        43.2%;   Score 495;  DB 1;  Length 247;
Best Local Similarity 41.3%;   Pred. No. 7. 81e-101;
Matches   66;  Conservative  43;  Mismatches 48;  Indels   3;  Gaps   3;

Db  88 INSAKVIRHPNYNSWILDNDIMLIKLSSPAVLNARVATISLPRACAAPGTQCLISGWGNT 147
       ::::  :iii::i : ii:iiiii  :  :iiii: : iiii::i  i:::i  ii:iiii
Qy   2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L  60

Db 148 LSSGTNYPELLQCLDAPILTQAQCEASYPCQITENMICAGFLEGGKDSCQGDSGGPVVCN 207
       i:i:i   i :iii:: :::::i   i      :i:iii  :   iii iiiiii::ii
Qy  61 LANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXXDSCNGDSGGPLICN 119

Db 208 GELQGIVSWGYG-CAQKNRPGVYTKVCNFVDWIQSTIAAN 246
       i:iiiiiiii:i  : iiiii i:iiii:i:i:  i:
Qy 120 GYLQLVSFGKAPCCQVGPGVGVYTNLCKFTEWIEKTVQAS 159


RESULT  15
ID   TRY1_BOVIN    STANDARD;        PRT;    243 AA.
AC   P00760;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN, CATIONIC PRECURSOR (EC 3.4.21.4) (FRAGMENT).
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RA   OKAJIMA T., MANIWA M., NAGAO S., FUJIKAWA H., GOTO S.;
RL   SUBMITTED (OCT-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE OF 15-243, AND DISULFIDE BONDS.
RX   MEDLINE; 67168848.
RA   MIKES O., HOLEYSOVSKY V., TOMASEK V., SORM F.;
RT   "Covalent structure of bovine trypsinogen. The position of the
RT   remaining amides.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 24:346-352(1966).
RN   [3]
RX   MEDLINE; 72035053.
RA   HARTLEY B.S.;
RT   "Homologies in serine proteinases.";
RL   PHILOS. TRANS. R. SOC. LOND., B, BIOL. SCI. 257:77-87(1970).
RN   [4]
RP   REVISIONS.
RX   MEDLINE; 75146445.
RA   TITANI K., ERICSSON L.H., NEURATH H., WALSH K.A.;
RT   "Amino acid sequence of dogfish trypsin.";
RL   BIOCHEMISTRY 14:1358-1366(1975).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF CALCIUM-BINDING SITE.
RX   MEDLINE; 76072097.
RA   BODE W., SCHWAGER P.;
RT   "The refined crystal structure of bovine beta-trypsin at 1.8-A
RT   resolution. II. Crystallographic refinement, calcium binding site,
RT   benzamidine binding site and active site at pH 7.0.";
RL   J. MOL. BIOL. 98:693-717(1975).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE; 77112431.
RA   KOSSIAKOFF A.A., CHAMBERS J.L., KAY L.M., STROUD R.M.;
RT   "Structure of bovine trypsinogen at 1.9-A resolution.";
RL   BIOCHEMISTRY 16:654-664(1977).
RN   [7]
RP   DISULFIDE BONDS.
RX   MEDLINE; 66079271.
```

```
RA   KAUFFMAN D.L.;
RT   "The disulphide bridges of trypsin.";
RL   J. MOL. BIOL. 12:929-932(1965).
CC   -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE ACINAR CELLS OF THE
CC       PANCREAS.
CC   -!- PTM: AUTOCATALYTIC CLEAVAGE AFTER LYS-20 LEADS TO BETA-TRYPSIN BY
CC       RELEASING A TERMINAL HEXAPEPTIDE. SUBSEQUENT CLEAVAGE AFTER
CC       LYS-145 LEADS TO ALPHA-TRYPSIN. FURTHER CLEAVAGE AFTER LYS-190
CC       YIELDS PSEUDOTRYPSIN. A CLEAVAGE MAY ALSO OCCUR AFTER ARG-119.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D38507; G559311; -.
DR   PIR; A00946; TRBOTR.
DR   PDB; 2TGA; 09-APR-85.
DR   PDB; 1TGB; 14-MAR-85.
DR   PDB; 1TGC; 09-APR-85.
DR   PDB; 2TGD; 07-MAY-86.
DR   PDB; 1TGN; 22-OCT-84.
DR   PDB; 2TGP; 14-MAR-85.
DR   PDB; 1TGS; 14-MAR-85.
DR   PDB; 1TGT; 15-OCT-90.
DR   PDB; 2TGT; 09-APR-85.
DR   PDB; 1TLD; 15-OCT-92.
DR   PDB; 2TLD; 15-JUL-92.
DR   PDB; 1TPA; 14-MAR-85.
DR   PDB; 2TPI; 15-JUL-91.
DR   PDB; 3TPI; 14-MAR-85.
DR   PDB; 4TPI; 08-NOV-85.
DR   PDB; 1TPO; 14-MAR-85.
DR   PDB; 1TPP; 16-APR-87.
DR   PDB; 1NTP; 15-APR-91.
DR   PDB; 3PTB; 14-MAR-85.
DR   PDB; 2PTC; 14-MAR-85.
DR   PDB; 2PTN; 15-APR-91.
DR   PDB; 3PTN; 09-APR-85.
DR   PDB; 5PTP; 07-JUL-97.
DR   PDB; 1TAB; 15-JUL-92.
DR   PDB; 1GBT; 31-JAN-94.
DR   PDB; 1PPC; 31-JAN-94.
DR   PDB; 1PPE; 31-JAN-94.
DR   PDB; 1PPH; 31-JUL-94.
DR   PDB; 1SMF; 31-JUL-94.
DR   PDB; 1TNG; 30-NOV-94.
DR   PDB; 1TNH; 30-NOV-94.
DR   PDB; 1TNI; 30-NOV-94.
DR   PDB; 1TNJ; 30-NOV-94.
DR   PDB; 1TNK; 30-NOV-94.
DR   PDB; 1TNL; 30-NOV-94.
DR   PDB; 1BTP; 29-JAN-96.
DR   PDB; 1BTW; 15-OCT-95.
DR   PDB; 1BTX; 15-OCT-95.
DR   PDB; 1BTY; 15-OCT-95.
DR   PDB; 1BTZ; 15-OCT-95.
DR   PDB; 1JRS; 14-OCT-96.
DR   PDB; 1JRT; 14-OCT-96.
DR   PDB; 1MAX; 14-OCT-96.
DR   PDB; 1MAY; 14-OCT-96.
DR   PDB; 1MTS; 20-AUG-97.
DR   PDB; 1MTU; 12-NOV-97.
DR   PDB; 1MTV; 12-NOV-97.
DR   PDB; 1MTW; 12-NOV-97.
DR   PDB; 1TPS; 26-JAN-95.
```

```
DR   PDB; 1TYN; 26-JAN-95.
DR   PDB; 1TAW; 24-JUN-97.
DR   PDB; 1AQ7; 25-FEB-98.
DR   PDB; 2TIO; 30-SEP-98.
DR   PDB; 1XUF; 16-DEC-98.
DR   PDB; 1XUG; 16-DEC-98.
DR   PDB; 1XUH; 11-NOV-98.
DR   PDB; 1XUI; 11-NOV-98.
DR   PDB; 1XUJ; 11-NOV-98.
DR   PDB; 1XUK; 11-NOV-98.
DR   PDB; 1BJU; 02-DEC-98.
DR   PDB; 1BJV; 02-DEC-98.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
DR   PFAM; PF00089; trypsin; 1.
KW   HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KW   3D-STRUCTURE.
FT   NON_TER         1       1
FT   SIGNAL         <1      14
FT   PROPEP         15      20
FT   CHAIN          21     243   TRYPSIN. CATIONIC (BETA-TRYPSIN).
FT   CHAIN          21     145   ALPHA-TRYPSIN CHAIN 1.
FT   CHAIN         146     243   ALPHA-TRYPSIN CHAIN 2.
FT   CA_BIND        72      82
FT   BINDING       191     192   SUBSTRATE.
FT   BINDING       194     195   SUBSTRATE.
FT   BINDING       197     197   SUBSTRATE.
FT   DISULFID       27     157
FT   DISULFID       45      61
FT   DISULFID      129     230
FT   DISULFID      136     203
FT   DISULFID      168     182
FT   DISULFID      193     217
FT   ACT_SITE       60      60   CHARGE RELAY SYSTEM.
FT   ACT_SITE      104     104   CHARGE RELAY SYSTEM.
FT   ACT_SITE      197     197   CHARGE RELAY SYSTEM.
FT   SITE          191     191   REQUIRED FOR SPECIFICITY.
FT   STRAND         22      22
FT   STRAND         25      26
FT   TURN           29      30
FT   TURN           34      34
FT   STRAND         35      39
FT   STRAND         43      51
FT   TURN           52      53
FT   STRAND         54      57
FT   HELIX          59      61
FT   STRAND         67      70
FT   TURN           75      76
FT   STRAND         83      92
FT   TURN           94      95
FT   STRAND         97      97
FT   TURN           98     101
FT   STRAND        102     102
FT   TURN          103     103
FT   STRAND        106     110
FT   STRAND        124     124
FT   TURN          132     133
FT   STRAND        135     140
FT   STRAND        156     162
FT   HELIX         165     171
FT   TURN          173     174
FT   TURN          178     179
FT   STRAND        180     183
FT   TURN          186     187
FT   STRAND        191     191
FT   TURN          194     195
FT   TURN          197     198
FT   STRAND        200     203
FT   TURN          204     205
FT   TURN          206     213
FT   STRAND        213     221
FT   TURN          224     228
FT   STRAND        229     241
FT   HELIX         241
FT   TURN          242     242
SQ   SEQUENCE      243 AA;  25425 MW;  E02FAF04 CRC32;
```

```
Query Match
Best Local Similarity  43.1%;   Score 494; DB 1; Length 243;
                       43.1%;   Pred. No. 1.46e-100;
Matches    69;  Conservative  36;  Mismatches  52;  Indels  3;  Gaps  3;

Db    85  ISASKSIVHPSYNSNTLNNDIMLIKLKSAASLNSRVASISLPTSCASAGTQCLISGWGNT  144
          |::||:||:||| | ||:|||||  : ::  ||::::::|::::|  ||:::|||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL   61

Db   145  KSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDSCQGDSGGPVVCS  204
          :::|||| ||:  ::::|     |      :|||||   ||| |||||||:::|:
Qy    62  -ANGRM-PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN  119

Db   205  GKLQGIVSWGSG-CAQKNKPGVYTKVCNYVSWIKQTIASN  243
          |:|||||::|:: |:|| || | |||:|:||:|:|::
Qy   120  GYIQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS  159

Search completed: Tue Sep 28 10:58:45 1999
Job time : 18 secs.
```

```
*********************************************
*   _   _  ____   ____  ____   ____ _   _   *
*  |  \/  ||  _ \ / ___||  _ \ / ___| | | |  *
*  | |\/| || |_) |\___ \| |_) | |   | |_| |  *
*  | |  | ||  __/  ___) |  _ <| |___|  _  |  *
*  |_|  |_||_|    |____/|_| \_\\____|_| |_|  (TM)
*                                           *
*********************************************

        Release 3.1A John F. Collins, Biocomputing Research Unit.
        Copyright (c) 1993-1998 University of Edinburgh, U.K.
               Distribution rights by Oxford Molecular Ltd

*********************************************

MPsrch_pp  protein - protein database search, using Smith-Waterman algorithm

Run on:        Tue Sep 28 10:59:02 1999;    MasPar time 13.07 Seconds
                                             663.951 Million cell updates/sec

Tabular output not generated.

Title:         >US-09-030-606-172
Description:   (1-159) from US09030606.pep
Perfect Score: 1145
Sequence:      1 MVEASLSVRHPEYNRPLLAN.........GVYTNLCKFTEWIEKTVQAS 159

Scoring table: PAM 150
               Gap 11

Searched:      179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
                 Listing first 45 summaries

Database:      sptrembl9
               1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
               5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
               9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
               13:sp_vertebrate 14:sp_virus

Statistics:    Mean 40.858;  Variance 64.854;  scale 0.630

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description	Pred. No.
1	507	44.3	249	13	Q92046	PREPROTRYPSIN PRECURSO	5.06e-97
2	503	43.9	250	13	Q93265	TRYPSINOGEN 1 PRECURSO	5.17e-96
3	480	41.9	242	13	Q93266	TRYPSINOGEN 2 PRECURSO	3.20e-90
4	471	41.1	282	4	O75837	SERINE PROTEASE (TLSP)	5.82e-88
5	469	41.0	260	11	Q61955	NEUROPSIN	1.29e-87
6	468	40.9	260	11	O88780	SERINE PROTEASE PRECUR	3.29e-87
7	464	40.5	261	11	O88309	PRORENIN-CONVERTING EN	3.31e-86
8	463	40.4	242	13	Q92099	TRYPSIN PRECURSOR (EC	5.89e-86
9	459	40.1	237	13	Q91515	TRYPSINOGEN (FRAGMENT)	5.91e-85
10	459	40.1	245	13	O42160	TRYPSINOGEN B2 PRECURS	5.91e-85
11	455	39.7	259	4	Q15665	TRYPSINOGEN IV B-FORM.	5.92e-84
12	453	39.6	239	11	Q63275	KALLIKREIN (FRAGMENT)	1.87e-83
13	453	39.6	244	13	O42159	TRYPSINOGEN B1 PRECURS	1.87e-83
14	452	39.5	260	4	O60259	NEUROPSIN PRECURSOR	3.33e-83
15	449	39.2	251	11	O54854	MYELENCEPHALON SPECIFI	1.87e-82
16	448	39.1	247	13	O42158	TRYPSINOGEN A2 PRECURS	3.32e-82
17	448	39.1	247	13	O42608	TRYPSINOGEN A3 PRECURS	3.32e-82
18	446	39.0	246	11	O88301	SERINE PROTEASE (BSP).	1.05e-81
19	437	38.2	235	11	Q63274	KALLIKREIN (FRAGMENT).	1.84e-79
20	422	36.9	261	11	Q61855	TISSUE KALLIKREIN.	9.90e-76
21	412	36.0	258	13	O13059	SERINE PROTEASE PRECUR	2.99e-73
22	409	35.7	258	13	Q91516	VENOM PLASMINOGEN ACTI	1.65e-72
23	406	35.5	261	6	Q29474	KALLIKREIN PRECURSOR (	9.11e-72
24	403	35.2	260	13	Q93502	SALMOBIN.	5.02e-71
25	402	35.1	260	13	O73800	SALMOBIN.	8.87e-71
26	400	34.9	135	11	Q62284	NERVE GROWTH FACTOR, G	2.77e-70
27	393	34.3	266	5	O27761	TRYPSIN PRECURSOR (EC	1.47e-68
28	386	33.7	257	13	O13069	KN-BJ2 PRECURSOR.	7.79e-67
29	384	33.5	264	5	O62561	TRYPSIN (EC 3.4.21.4)	2.42e-66
30	382	33.4	257	13	O13062	SERINE PROTEASE PRECUR	7.49e-66
31	378	33.0	258	13	O42207	CAPILLARY PERMEABILITY	7.18e-65
32	377	32.9	257	13	Q91509	MUCOFIRASE 3.	1.26e-64
33	374	32.7	219	13	Q91036	TRYPSINOGEN I (FRAGMEN	6.87e-64
34	374	32.7	257	13	Q91508	MUCOFIRASE 2.	6.87e-64
35	374	32.7	258	13	O13063	SERINE PROTEASE PRECUR	6.87e-64
36	374	32.7	260	13	O13060	SERINE PROTEASE PRECUR	6.87e-64
37	373	32.6	257	13	Q91510	MUCOFIRASE 4.	1.21e-63
38	372	32.5	257	13	Q91511	MUCOFIRASE 5.	2.12e-63
39	372	32.5	257	13	Q91500	PREPROTRIMUBIN PRECURS	2.12e-63
40	368	32.1	325	5	O15944	26KDA PROTEASE.	2.02e-62
41	365	31.9	260	13	O13061	SERINE PROTEASE PRECUR	1.09e-61
42	364	31.8	257	13	Q91507	MUCOFIRASE 1.	1.92e-61
43	363	31.7	260	13	O13057	SERINE PROTEASE PRECUR	3.37e-61
44	359	31.4	250	11	O03955	GLANDULAR KALLIKREIN (	3.19e-60
45	357	31.2	178	13	O93594	TRYPSIN (EC 3.4.21.4)	9.79e-60

ALIGNMENTS

```
RESULT   1
ID       Q92046             PRELIMINARY;      PRT;     249 AA.
AC       Q92046; O13261;
DT       01-NOV-1996 (TREMBLREL.  01, CREATED)
DT       01-NOV-1996 (TREMBLREL.  01, LAST SEQUENCE UPDATE)
DT       01-NOV-1998 (TREMBLREL.  08, LAST ANNOTATION UPDATE)
DE       PREPROTRYPSIN PRECURSOR (EC 3.4.21.4) (ALPHA- AND BETA-TRYPSIN).
OS       DISSOSTICHUS MAWSONI.
OC       EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC       TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC       NOTOTHENIOIDEI; NOTOTHENIIDAE; DISSOSTICHUS.
RN       [1]
RP       SEQUENCE FROM N.A.
RC       TISSUE=PANCREAS;
RA       CHEN L., DEVRIES A.L., CHENG C.H.C.;
RL       SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN       [2]
RP       SEQUENCE FROM N.A.
RX       MEDLINE; 97268652.
RA       CHEN L., DEVRIES A.L., CHENG C.H.;
RT       "Evolution of antifreeze glycoprotein gene from a trypsinogen gene in
RT       Antarctic notothenioid fish.";
RL       PROC. NATL. ACAD. SCI. U.S.A. 94:3811-3816(1997).
CC       -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-|-XAA, LYS-|-XAA.
DR       EMBL; U58945; G1399811; -.
DR       EMBL; U58835; G2102686; -.
DR       PFAM; PF00089; trypsin; 1.
KW       SIGNAL; HYDROLASE.
FT       SIGNAL        1     15                      POTENTIAL.
FT       CHAIN        16    249                      PROTRYPSIN.
SQ       SEQUENCE   249 AA;   27169 MW;   55C59EB4 CRC32;

Query Match          44.3%;   Score 507;  DB 13;  Length 249;
Best Local Similarity 42.0%;  Pred. No. 5.06e-97;
Matches  68;  Conservative  48;  Mismatches  42;  Indels  4;  Gaps  3;

Db    88  LMKTNTIIWHPSYDYQTLDFPDIMLIKLYHPVEVTEAVAPLPTSCPYGGLSCSVSGWGI  147
          ::  :: ||  : :|| | | ||||||   :|:  |   :   || :| ||||| ||||||
Qy     1  MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDIIRSISIASQCPTAGNSCLVSGWGL   60

Db   148  AKLGGEAYMPTLLQCLNVPIVDQQVCENTYPGLISTTMVCAGYMEGGKDACNGDSGSPLV  207
          :||||| :|  :  | || :  :  || ||||||||  ::: |||||||||||||||||| |
Qy    61  --LANGR-MPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLI  117
```

```
Best Local Similarity 41.1%;    Pred. No. 3.20e-90;    Indels   2;   Gaps   2;
Matches  65;  Conservative  45;  Mismatches  46;

Db   85 VSSSRVIRHPNYDSWNIDNDIMLIKLSKPATLNQYVKTVALPSSCAPAGTMCKVSGWGNT 144
          :|:  ::|||:| :: : :::::|||||| : :: :::::::|||| :|:|||||||||| ||||
Qy    2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L  60

Db  145 MSSADNGDLLQCLDIPILSFSDCNNAYPGMITDSMFCAGYLEGGKDSCQGDSGGPVICNG 204
          :::: :||:|::::::|| |: :  |:|| |  |||||  ||| ||||||||||:|||
Qy   61 LANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db  205 ELQGVVSWGYG-CAERGNPGVYAKVCLFNDWLESTMAS 241
          |||:|: |: ||::||:|::| :|
Qy  121 YLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA  158
```

RESULT   4
```
ID   075837                 PRELIMINARY;         PRT;      282 AA.
AC   075837;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   SERINE PROTEASE (TLSP).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98438738.
RA   YOSHIDA S., TANIGUCHI M., SUEMOTO T., OKA T., HE X.P., SHIOSAKA S.;
RT   "cDNA cloning and expression of a novel serine protease, TLSP.";
RL   BIOCHIM. BIOPHYS. ACTA 1399:225-228(1998).
DR   EMBL; AB012917; D1034370; -.
KW   PROTEASE; SERINE PROTEASE.
FT   CHAIN        54    282       TLSP.
SQ   SEQUENCE   282 AA;   31059 MW;   E97E294C CRC32;

Query Match             41.1%;    Score 471;   DB 4;   Length 282;
Best Local Similarity   41.5%;    Pred. No. 5.82e-88;   Indels  2;   Gaps  1;
Matches  59;  Conservative  38;  Mismatches  43;

Db  141 NDIMLVKMASPVSITWAVRPLTLSSRCVTAGTSCLISGWGSTSSPQLRLPHTLRCANITI 200
          ||:||:|| :|| :|| ||| |||:|||||||||||:||::|||:||||||||||||||
Qy   20 NDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANG--RMPTVLQCVNVSV  77

Db  201 IEHQKCENAYPGNITDTMVCASVQEGGKDSCQGDSGGPLVCNQSLQGLISWGQDPCAITR 260
          : :  :|: ::|    :|| |||||| |||||||||||:||  |||:|:||||||:|::
Qy   78 VSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLVSFGKAPCGQVG 137

Db  261 KPGVYTKVCKYVDWIQETMKNN 282
          ||||||:|| ::|:||:|:
Qy  138 VPGVYTNLCKFTEWIEKTVQAS 159
```

RESULT   5
```
ID   Q61955                 PRELIMINARY;         PRT;      260 AA.
AC   Q61955;
DT   01-NOV-1996 (TREMBLREL. 01, CREATED)
DT   01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT   01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE   NEUROPSIN.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=HIPPOCAMPUS;
RX   MEDLINE; 95348817.
RA   CHEN Z.-L., YOSHIDA S., KATO K., MOMOTA Y., SUZUKI J., TANAKA T.,
RA   ITO J., NISHINO H., AIMOTO S., KIYAMA H., SHIOSAKA S.;
RT   "Expression and activity-dependent changes of a novel limbic-serine
RT   protease gene in the hippocampus.";
RL   J. NEUROSCI. 15:5088-5097(1995).
```

```
Db  208 CDGEVQGLVSWGQG-CAEPNYPGVYVKLCEFHSWEEVLAAN 248
          |:|:||||| : |: |||  ::|:|| ||| :|| | |: :
Qy  118 CNGYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159
```

RESULT   2
```
ID   093265                 PRELIMINARY;         PRT;      250 AA.
AC   093265;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN 1 PRECURSOR (EC 3.4.21.4).
GN   TRP1.
OS   PLEURONECTES AMERICANUS.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC   PLEURONECTIFORMES; PLEURONECTOIDEI; PLEURONECTIDAE; PLEURONECTES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PANCREAS;
RA   DOUGLAS S.E., GALLANT J.W.;
RT   "Isolation of cDNAs for Trypsinogen from the Winter Flounder,
RT   Pleuronectes americanus.";
RL   J. MAR. BIOTECHNOL. 0:0-0(1998).
DR   EMBL; AF012462; G3452117; -.
KW   SIGNAL; HYDROLASE.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        23    250       TRYPSINOGEN 1.
SQ   SEQUENCE   250 AA;   27466 MW;   D387394D CRC32;

Query Match             43.9%;    Score 503;   DB 13;   Length 250;
Best Local Similarity   44.4%;    Pred. No. 5.17e-96;   Indels  5;   Gaps  3;
Matches  72;  Conservative  36;  Mismatches  49;

Db   88 LMKTDTIIWHPNYDYQTLDFDIMLIKLYHPVEVTEAVAPISLPSSCPVGTPCSVSGWGN 147
          :|:  :|| ::|: : |:|  :||||| :   ||  : :: ||| :|| :| ||||||
Qy    1 MVEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGL  60

Db  148 TARDGDDVYMPTLLQCMDVPIIDEEQCMKSYPGMISPRMVCAGFMDGSRDACNGDSGSPL 207
          :| | : : |  |:|  :| : :||  |   ||| :|:|||| :| |: ||||||||:||
Qy   61 LA-NGR--MPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPL  116

Db  208 VCRGEVTGLVSWGGQ-CAQPNYPGVYVKLCEFHAWIENTLAA 248
          ||| |  |||| | | ||  :|:|||| ||| |  |  | |
Qy  117 ICNGYLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA  158
```

RESULT   3
```
ID   093266                 PRELIMINARY;         PRT;      242 AA.
AC   093266;
DT   01-NOV-1998 (TREMBLREL. 08, CREATED)
DT   01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN 2 PRECURSOR (EC 3.4.21.4).
GN   TRP2.
OS   PLEURONECTES AMERICANUS.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC   PLEURONECTIFORMES; PLEURONECTOIDEI; PLEURONECTIDAE; PLEURONECTES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=INTESTINE, PYLORIC CAECA, AND PANCREAS;
RA   DOUGLAS S.E., GALLANT J.W.;
RT   "Isolation of cDNAs for Trypsinogen from the Winter Flounder,
RT   Pleuronectes americanus.";
RL   J. MAR. BIOTECHNOL. 0:0-0(1998).
DR   EMBL; AF012463; G3452120; -.
KW   SIGNAL; HYDROLASE.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        21    242       TRYPSINOGEN 2.
SQ   SEQUENCE   242 AA;   26180 MW;   7E72C91C CRC32;

Query Match             41.9%;    Score 480;   DB 13;   Length 242;
```

```
DR   EMBL; D30785; D1007022; -.
DR   MGD; MGI:1202387; NRPN.
DR   PFAM; PF00089; trypsin; 1.
KW   SERINE PROTEASE.
SQ   SEQUENCE   260 AA;  28523 MW;  232F78DB CRC32;

Query Match             41.0%;  Score 469;  DB 11;  Length 260;
Best Local Similarity   39.1%;  Pred. No. 1.85e-87;
Matches   63;  Conservative   49;  Mismatches   43;  Indels    6;  Gaps    4;

Db     98  IQVAQSIQHPCYNNSNPEDHSHDIMLIRLQNSANLGDKVKPVQLANLCPKVGQKCIISGW  157
               ::::|||| |   :|||:::|   :||::: :| ::::|  || ||:|||
Qy      2  VEASLSVRHPEYNR--PL-LANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGW   58

Db    158  GTVTSPQENFPNTLNCAEVKIYSQNKCERAYPGKITEGMVCAGSSNG-ADTCQGDSGGPL  216
               |  |||||||  | | :: ::|:||| ||||: ||||:||| :| ||| ||||||||
Qy     59  GLLANGR--MPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPL  116

Db    217  VCDGMLQGITSWGSDPCGKPEKPGVYTKICRYTWIKKTMD  257
               ||  :|||:||||  || ||| |||| | || | |::|
Qy    117  ICNGYLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQ  157


RESULT   6
ID   O88780              PRELIMINARY;       PRT;   260 AA.
AC   O88780;
DT   01-NOV-1998 (TREMBLREL.  08, CREATED)
DT   01-NOV-1998 (TREMBLREL.  08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL.  08, LAST ANNOTATION UPDATE)
DE   SERINE PROTEASE PRECURSOR.
GN   BSP1.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=BRAIN;
RX   MEDLINE; 98389725.
RA   DAVIES B.J., PICKARD B.S., STEEL M., MORRIS R.G., LATHE R.;
RT   "Serine Proteases in Rodent Hippocampus.";
RL   J. BIOL. CHEM. 273:23004-23011(1998).
DR   EMBL; AJ005641; E1319201; -.
KW   SIGNAL; PROTEASE; SERINE PROTEASE.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    260       SERINE PROTEASE.
SQ   SEQUENCE   260 AA;  28510 MW;  50894868 CRC32;

Query Match             40.9%;  Score 468;  DB 11;  Length 260;
Best Local Similarity   41.4%;  Pred. No. 3.29e-87;
Matches   58;  Conservative   37;  Mismatches   42;  Indels    3;  Gaps    2;

Db    118  SHDIMLIRLQNSANLGDKVKPIELANLCPKVGQKCIISGWGTVTSPQENFPNTLNCAEVK  177
               :|||||:||:::|||||| || :|||| | |||||: ::|:| ::::
Qy     19  ANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGR--MPTVLQCVNVS   76

Db    178  IYSQNKCERAYPGKITEGMVCAGSSNG-ADTCQGDSGGPLVCNGVLQGITTWGSDPCGKP  236
               ::|:||| ||||: ||||:||| :| ||| ||||||||:|||| |||| :: |   |
Qy     77  VVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLVSFGKAPCGQV  136

Db    237  EKPGVYTKICRYTNWIKKTM  256
               ||| |||| ||| | |::|
Qy    137  GVPGVYTNLCKFTEWIEKTV  156


RESULT   7
ID   O88309              PRELIMINARY;       PRT;   261 AA.
AC   O88309;
DT   01-NOV-1998 (TREMBLREL.  08, CREATED)
DT   01-NOV-1998 (TREMBLREL.  08, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL.  08, LAST ANNOTATION UPDATE)
DE   PRORENIN-CONVERTING ENZME (MK13B) PRECURSOR (MK13B).
GN   MKLK-13B.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98351995.
RA   HOSOI K., TADA J., KANAMORI N., YAMANAKA N.;
RT   "Expression of an allozyme of prorenin-converting enzyme in the
RT   submandibular gland of DBA/2N mice.";
RL   J. BIOCHEM. 124:368-376(1998).
DR   EMBL; AB016032; D1032647; -.
KW   SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
SQ   SEQUENCE   261 AA;  28692 MW;  8DB1814F CRC32;

Query Match             40.5%;  Score 464;  DB 11;  Length 261;
Best Local Similarity   43.3%;  Pred. No. 3.31e-86;
Matches   61;  Conservative   34;  Mismatches   44;  Indels    2;  Gaps    1;

Db    117  FSNDLMLRLSKPADITDVVKPIALPTKEPKPGSKCLASGWGSITPTRWQKPDDLQCVFI  176
               ::|||||::|   -  ::::|::: :||||: :::|: ||||  ::: ||||
Qy     18  LANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGR--MPTVLQCVNV   75

Db    177  TLLPNENCAKVYLQKVTDVMLCAGEMGGGKDTCRDDSGGPLICDGILQGTTSYGPTPCGK  236
               ::::|:|:|:|   |:||   :|||  | ||||||||||:|||: ||:|: :|||
Qy     76  SVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLVSFGKAPCGQ  135

Db    237  PGVPAIYTNLIKFNSWIKDTM  257
               ||::|||||:|| |:|
Qy    136  VGVPGVYTNLCKFTEWIEKTV  156


RESULT   8
ID   Q92099              PRELIMINARY;       PRT;   242 AA.
AC   Q92099;
DT   01-NOV-1996 (TREMBLREL.  01, CREATED)
DT   01-NOV-1996 (TREMBLREL.  01, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL.  08, LAST ANNOTATION UPDATE)
DE   TRYPSIN PRECURSOR (EC 3.4.21.4) (ALPHA- AND BETA-TRYPSIN).
OS   PARANOTOTHENIA MAGELLANICA.
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC   TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA; PERCIFORMES;
OC   NOTOTHENIOIDEI; NOTOTHENIIDAE; PARANOTOTHENIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PYLORIC CAECA;
RX   MEDLINE; 97104330.
RA   GENICOT S., RENTIER-DELRUE F., EDWARDS D., VAN BEEUMEN J., GERDAY C.;
RT   "Trypsin and trypsinogen from an Antarctic fish: molecular basis of
RT   cold adaptation.";
RL   BIOCHIM. BIOPHYS. ACTA 1298:45-57(1996).
DR   EMBL; X82223; G559508; -.
DR   PFAM; PF00089; trypsin; 1.
KW   SIGNAL; HYDROLASE.
FT   SIGNAL        1     13       POTENTIAL.
FT   CHAIN        14    242       PROTRYPSIN.
SQ   SEQUENCE   242 AA;  26201 MW;  DDD27DCA CRC32;

Query Match             40.4%;  Score 463;  DB 13;  Length 242;
Best Local Similarity   39.7%;  Pred. No. 5.89e-86;
Matches   62;  Conservative   43;  Mismatches   49;  Indels    2;  Gaps    2;

Db     85  ISSSRVIRHPNYSSYNIDNDIMLIKLKPATLNQYVQAVALPSSCAPAGTMCTVSGWGST  144
               ::: | ||:||     ::::::|||| :: : ||  |  ||| ||: |||||
Qy      2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL   61

Db    145  QSSSADGNKLQCLNIPILSDRDCDNSYPGMITDAMFCAGYLQGGKDSCQGDSGGPVVCNG  204
               ::: | |    |::|::     |:  :|| |  |||| |||  ::|| |||||:|||
Qy     62  ANGRMPTV-LQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG  120

Db    205  ELQGVVSWGYG-CAERDHPGYAKVCLFNDWLETSM  239
               ||| | ||| |   |  || :   ||:   | |
Qy    121  YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTV  156
```

```
RESULT    9
ID    Q91515
AC    Q91515;
DT    01-NOV-1996 (TREMBLREL. 01, CREATED)
DT    01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT    01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE    TRYPSINOGEN (FRAGMENT).
OS    FUGU RUBRIPES (JAPANESE PUFFERFISH).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;
OC    TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;
OC    TETRAODONTIFORMES; TETRAODONTOIDEI; TETRAODONTIDAE; FUGU.
RN    [1]
RP    SEQUENCE FROM N.A.
RA    WANG K., GAN L., LEE I., ROACH J., HOOD L.;
RL    SUBMITTED (APR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; U25747; G971196; -.
DR    PFAM; PF00089; trypsin; 1.
FT    NON_TER       1      1
SQ    SEQUENCE    237 AA;   25726 MW;  6A296989 CRC32;

Query Match             40.1%;  Score 459;  DB 13;  Length 237;
Best Local Similarity   38.6%;  Pred. No. 5.91e-85;
Matches   61; Conservative   46; Mismatches   49; Indels    2; Gaps    2;

Db    80  ISSSRVIRHPNYSSYNIDNDIMLIKLSKPATLNQYVQPVALPSSCAAAGTMCKVSGWGNT 139
              :: :|||||| : |:|||| | |||::::::|| |:||||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db   140  MSSTADRNKLQCLNIPILSDRDCENSYPGMITDAMFCAGYLEGGKDSCQGDSGGPVVCNN 199
              : :                  ||| | :  |||||  |||| |||||:||
Qy    61  LANGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db   200  ELQGVVSWGYG-CAERDHPGVYAKVCLFNDWLESTMAS 236
              ||||||:|| |   |||: :|:::: |   :::||
Qy   121  YLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQA 158
```

```
Db   208  QLQGIVSWGRG-CALPNPGVYTKVCNYNSWIASTMAAN 245
              ||||||:|| |   ||||:|||::|:::: |:  :|:
Qy   121  YLQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT   11
ID    Q15665
AC    Q15665;
DT    01-NOV-1996 (TREMBLREL. 01, CREATED)
DT    01-JAN-1999 (TREMBLREL. 09, LAST SEQUENCE UPDATE)
DT    01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE    TRYPSINOGEN IV B-FORM.
OS    HOMO SAPIENS (HUMAN).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC    CATARRHINI; HOMINIDAE; HOMO.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=BRAIN;
RX    MEDLINE; 94123994.
RA    WIEGAND U., CORBACH S., MINN A., KANG J., MUELLER-HILL B.;
RT    "Cloning of the cDNA encoding human brain trypsinogen and
RT    characterization of its product.";
RL    GENE 136:167-175(1993).
DR    EMBL; X71345; E1358280; -.
SQ    SEQUENCE    259 AA;   28001 MW;  C6FE2AF0 CRC32;

Query Match             39.7%;  Score 455;  DB 4;   Length 259;
Best Local Similarity   41.3%;  Pred. No. 5.92e-84;
Matches   57; Conservative   34; Mismatches   57; Indels    3; Gaps    3;

Db   100  INAAKIIRHPKYNRDTLDNDIMLIKLSSPAVINARVSTISLPTAPPAAGTECLISGWGNT 159
              :|: |||| ||  :| |||:|| | ||:||||:::|||||  |:|| ||::|||:|||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db   160  LSFGADYPDELKCLDAPVLTQAECKASYPGKITNSMFCVGFLEGGKDSCQRDSGGPVVCN 219
              |::| :       :|:: :|::  |   |:||:|:|||| || |||||:||
Qy    61  LANGRM-PTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICN 119

Db   220  GQLQGVVSWGHG-CAWKNRPGVYTKVNYVDWIKDTIAAN 258
              | |||||:|| |   ||:|||||:|: ::: || : :|
Qy   120  GYILQGLVSFGKAPCGQVGPGVYTNLCKFTEWIEKTVQAS 159

RESULT   12
ID    Q63275
AC    Q63275;
DT    01-NOV-1996 (TREMBLREL. 01, CREATED)
DT    01-NOV-1998 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT    01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE    KALLIKREIN (FRAGMENT).
OS    RATTUS NORVEGICUS (RAT).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC    SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
RA    ZINT2 C.B., MA J.X., CHAO J., CHAO L.;
RL    SUBMITTED (JAN-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; L33840; G609585; -.
FT    PFAM; PF00089; trypsin; 1.
FT    NON_TER       1      1
SQ    SEQUENCE    239 AA;   26382 MW;  0609E3E6 CRC32;

Query Match             39.6%;  Score 453;  DB 11;  Length 239;
Best Local Similarity   46.4%;  Pred. No. 1.87e-83;
Matches   46; Conservative   26; Mismatches   46; Indels    3; Gaps    3;

Db    96  SNDLMLHLSEPADITDGVKIVIDLPTKEPKVGSTCLVSGWGSTNPSEWFPDDLQCVNIH 155
              :||:|| | :      :||:| |||:  | |  || ||||:|  ||:
Qy    19  ANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLL-ANGR-MPTVLQCVNVS 76

Db   156  LLSNEKCIKAYKEKVTDLMLCAGELEGGKDTCRGDSGGPLICDGVLQGITSWGSVPCGEP 215
              ::||| | | :|  || :|||||| |||| | |||| ||:| :|||||| |  ||:|
Qy    61  LANGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120
```

```
RESULT   10
ID    Q42160
AC    Q42160;
DT    01-JAN-1998 (TREMBLREL. 05, CREATED)
DT    01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT    01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE    TRYPSINOGEN B2 PRECURSOR (FRAGMENT).
GN    TRYPB2.
OS    PETROMYZON MARINUS (SEA LAMPREY).
OC    EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; CEPHALASPIDOMORPHI;
OC    PETROMYZONTIFORMES; PETROMYZONTIDAE; PETROMYZON.
RN    [1]
RP    SEQUENCE FROM N.A.
RA    ROACH J.C.;
RL    SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR    EMBL; AF011901; G2367501; -.
DR    PFAM; PF00089; trypsin; 1.
KW    SIGNAL.
FT    NON_TER       1      1
FT    SIGNAL       <1     13          POTENTIAL.
FT    CHAIN        14    245          TRYPSIN B2.
SQ    SEQUENCE    245 AA;   26001 MW;  071872F0 CRC32;

Query Match             40.1%;  Score 459;  DB 13;  Length 245;
Best Local Similarity   40.3%;  Pred. No. 5.91e-85;
Matches   64; Conservative   40; Mismatches   53; Indels    2; Gaps    2;

Db    88  IQASKAIRHPQYNSATIDNDIMLIKLSSPATLNQYAQAIPLPSSCVGTGVMCTISGWGET 147
              :||: ||| || |   :||||||| | | | || | : |: | |||| |  | |||
Qy     2  VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db   148  QTSVGSPDVLMCVQAPVLSDTSCRNSYPGDITNNMICLGYLEGGKDSCQGDSGGPVVCNG 207
              |||| |   | | :   | | |::|| |||:| | || |||| |||||:|||
Qy    61  LANGRMPTVLQCVNVSVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120
```

```
Qy  77 VVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLVSFGKAPCGQV 136
Db 216 NKPGIYTKLIKFTSWI-KEV 234
Qy 137 GVPGVYTNLCKFTEWIEKTV 156

RESULT 13
ID   042159        PRELIMINARY;       PRT;     244 AA.
AC   042159;
DT   01-JAN-1998 (TREMBLREL. 05, CREATED)
DT   01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   TRYPSINOGEN B1 PRECURSOR (FRAGMENT).
GN   TRYPB1.
OS   PETROMYZON MARINUS (SEA LAMPREY).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA: CEPHALASPIDOMORPHI;
OC   PETROMYZONTIFORMES; PETROMYZONTIDAE; PETROMYZON.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROACH J.C.;
RL   SUBMITTED (JUL-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR   EMBL; AF011900; G2367499; -.
DR   PFAM; PF00089; trypsin; 1.
KW   SIGNAL.
FT   NON_TER        1        1
FT   SIGNAL        <1       12        POTENTIAL.
FT   CHAIN         13      244        TRYPSIN B1.
SQ   SEQUENCE   244 AA;  25903 MW;  41F78768 CRC32;

Query Match             39.6%;  Score 453;  DB 13;  Length 244;
Best Local Similarity   39.0%;  Pred. No.1.87e-83;
Matches    62;  Conservative  42;  Mismatches 53;  Indels  2;  Gaps  2;

Db  87 IQASKAIRHPQYSSATIDNDIMLIKLSSPATLNQYAQAVPLPSSCVGTGVMCTISGWGET 146
Qy   2 VEASLSVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWG-L 60

Db 147 QTSVGSPDVLMCVQAPVLSDTSCRNSYPGDITNNMICLGYLEGGKDSCQGDSGGPVVCNG 206
Qy  61 LANGRMPTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNG 120

Db 207 QLQGIVSWGRG-CALPNYPGVYTKVCNYNSWIASTMAAN 244
Qy 121 YLQGLVSFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQAS 159

RESULT 14
ID   060259        PRELIMINARY;       PRT;     260 AA.
AC   060259;
DT   01-AUG-1998 (TREMBLREL. 07, CREATED)
DT   01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   NEUROPSIN PRECURSOR.
GN   HNP.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC   CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HIPPOCAMPUS;
RX   MEDLINE; 98372070.
RA   YOSHIDA S., KITAYOSHI H., KUWAE K., INOUE N., SHIOSAKA S.;
RT   "Sequence analysis and expression of human neuropsin cDNA and gene.";
RL   GENE 213:9-16(1998).
DR   EMBL; AB009849; D1029613; -.
DR   EMBL; AB012761; D1029616; -.
KW   SIGNAL.
FT   SIGNAL         1       28        POTENTIAL.
FT   CHAIN         33      260        NEUROPSIN.
SQ   SEQUENCE   260 AA;  28048 MW;  0FAAA451 CRC32;

Query Match             39.5%;  Score 452;  DB 4;  Length 260;
```

```
Best Local Similarity   41.0%;   Pred. No. 3.33e-83;
Matches    57;  Conservative  30;  Mismatches 49;  Indels  3;  Gaps  2;

Db 119 HDLMLLQLRDQASLGSKVKPISLADHCTQPGQKCTVSGWGTVTSPRENFPDTLNCAEVKI 178
Qy  20 NDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGR--MPTVLQCVNVSV 77

Db 179 FPQKKCEDAYPGQITDGMVCAGSSKG-ADTCQGDSGGPLVCDGALQGITSWGSDPCGRSD 237
Qy  78 VSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLVSFGKAPCGQVG 137

Db 238 KPGVYTNICRYLDWIKKII 256
Qy 138 VPGVYTNLCKFTEWIEKTV 156

RESULT 15
ID   054854        PRELIMINARY;       PRT;     251 AA.
AC   054854;
DT   01-JUN-1998 (TREMBLREL. 06, CREATED)
DT   01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT   01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE   MYELENCEPHALON SPECIFIC PROTEASE.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC   SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 97477435.
RA   SCARISBRICK I.A., TOWNER M.D., ISACKSON P.J.;
RT   "Nervous system-specific expression of a novel serine protease:
RT   regulation in the adult rat spinal cord by excitotoxic injury.";
RL   J. NEUROSCI. 17:8156-8168(1997).
DR   EMBL; AF016269; G2853366; -.
KW   PROTEASE.
SQ   SEQUENCE   251 AA;  28014 MW;  21D0D8A9 CRC32;

Query Match             39.2%;  Score 449;  DB 11;  Length 251;
Best Local Similarity   39.7%;  Pred. No.1.87e-82;
Matches    60;  Conservative  30;  Mismatches 61;  Indels  0;  Gaps  0;

Db  99 TIVHPRYNPQTHDNDIMVHLKRPVKFSQRIQPLPLKKDCSEKNPDCQILGWGKMENGEF 158
Qy   7 SVRHPEYNRPLLANDLMLIKLDESVSESDTIRSISIASQCPTAGNSCLVSGWGLLANGRM 66

Db 159 PDTIQCADVQLVSREECERAYPGKITRSMVCAGDKREGNDSCQGDSGGPLVCGGHLRGIV 218
Qy  67 PTVLQCVNVSVVSEEVCSKLYDPLYHPSMFCAGGGQXQXDSCNGDSGGPLICNGYLQGLV 126

Db 219 SWGDMPCGSKEKPGVYTDVCTHIRWIQNIIR 249
Qy 127 SFGKAPCGQVGVPGVYTNLCKFTEWIEKTVQ 157

Search completed: Tue Sep 28 10:59:59 1999
Job time : 57 secs.
```

This Page Blank (uspto)

OM of: US-09-030-606-110 to: A_Geneseq_36:*   out_format : pfs
Date: Sep 25, 1999 11:34 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/USPTO30606/runat_24091999_171616_29804/app_query.fasta.1
-DB=A_Geneseq_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.000 -XGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-110
Query length: 3410
Database: A_Geneseq_36:*
Database sequences: 188963
Database length: 2368106
Search time (sec): 185.540000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:W69385	+	2861.00	3427.14	3.6e-184	553	! Prostate tumour specific gene c
A_Geneseq_36:W71869	+	2861.00	3427.14	3.6e-184	553	! Amino acid encoded by prostate
A_Geneseq_36:W85068	+	1287.00	1542.17	7.7e-79	255	! Amino acid encoded by the consensu
A_Geneseq_36:R80327	-	395.00	459.58	5.2e-19	762	! Protein polymeric adhesion subs
A_Geneseq_36:W57668	-	395.00	459.58	5.2e-19	762	! Collagen-like polymer. Recombin
A_Geneseq_36:W49715	-	395.00	459.58	5.2e-19	762	! Protein polymeric adhesion substi
A_Geneseq_36:R80324	-	391.00	454.77	9.6e-19	762	! Protein polymeric adhesion subs
A_Geneseq_36:W53535	-	391.00	454.77	9.6e-19	762	! Amino acid sequence of the plas
A_Geneseq_36:W57666	-	391.00	454.77	9.6e-19	762	! Collagen-like polymer. Recombin
A_Geneseq_36:W49713	-	391.00	454.77	9.6e-19	762	! Protein polymer adhesive substi
A_Geneseq_36:R80330	-	370.50	431.16	2.2e-17	682	! Protein polymeric adhesion subs
A_Geneseq_36:W57670	-	370.50	431.16	2.2e-17	682	! Collagen-like polymer. Recombin
A_Geneseq_36:W49717	-	370.50	431.16	2.2e-17	682	! Protein polymer adhesive substi
A_Geneseq_36:R28916	-	368.00	422.89	3.6e-17	1196	! Type III procollagen (prior ar
A_Geneseq_36:R80324	+	365.00	423.51	5.3e-17	762	! Protein polymeric adhesion subs
A_Geneseq_36:W53535	+	365.00	423.51	5.3e-17	762	! Amino acid sequence of the plas
A_Geneseq_36:W57666	+	365.00	423.51	5.3e-17	762	! Collagen-like polymer. Recombin
A_Geneseq_36:W49713	+	365.00	423.51	5.3e-17	762	! Protein polymer adhesive substi
A_Geneseq_36:R80327	+	364.00	422.31	6.2e-17	762	! Protein polymeric adhesion subs
A_Geneseq_36:W57668	+	364.00	422.31	6.2e-17	762	! Collagen-like polymer. Recombin
A_Geneseq_36:W49715	+	364.00	422.31	6.2e-17	762	! Protein polymer adhesive substi
A_Geneseq_36:R93254	-	360.00	414.37	1.2e-16	1064	! Collagen-like polymer sequence
A_Geneseq_36:W57652	-	360.00	414.37	1.2e-16	1064	! Collagen-like polymer. Recombi
A_Geneseq_36:R37704	-	360.00	414.36	1.2e-16	1065	! Collagen-like polymer DCP3 enc
A_Geneseq_36:R71704	-	357.50	411.24	1.8e-16	1078	! Collagen alpha 1 (III) chain F
A_Geneseq_36:R80334	-	357.00	413.10	1.8e-16	829	! Protein polymeric adhesion subs
A_Geneseq_36:W57673	-	357.00	413.10	1.8e-16	829	! Collagen-like polymer. Recombin
A_Geneseq_36:W49723	-	357.00	413.10	1.8e-16	829	! Protein polymer adhesive substi
A_Geneseq_36:R80320	-	356.50	412.41	2.0e-16	837	! CLP 3.7 protein multimer. Prote
A_Geneseq_36:W26353	-	356.50	412.41	2.0e-16	837	! Collagen-like protein CLP 3.7 (
A_Geneseq_36:W55530	-	356.50	412.41	2.0e-16	837	! Amino acid sequence for the CLP
A_Geneseq_36:W57659	-	356.50	412.41	2.0e-16	837	! Collagen-like polymer. Recombin
A_Geneseq_36:R37745	-	353.50	406.55	3.3e-16	1065	! Collagen-like polymer DCP4 enc
A_Geneseq_36:R93256	-	353.50	406.55	3.3e-16	1065	! Collagen-like polymer sequence
A_Geneseq_36:W57654	-	353.50	406.55	3.3e-16	1065	! Collagen-like polymer. Recombi
A_Geneseq_36:W68485	-	347.00	395.75	9.7e-16	1464	! Human recombinant collagen pro
A_Geneseq_36:R37746	-	344.50	400.61	1.2e-15	633	! Collagen-like polymer DCP5 enc
A_Geneseq_36:R93257	-	344.50	400.61	1.2e-15	633	! Collagen-like polymer sequence
A_Geneseq_36:W57655	-	344.50	400.61	1.2e-15	633	! Collagen-like polymer. Recombin
A_Geneseq_36:R53257	-	337.50	382.20	4.4e-15	1838	! Human collagen (Type V). Human
A_Geneseq_36:R93254	+	335.00	384.32	5.8e-15	1064	! Collagen-like polymer sequence
A_Geneseq_36:W57652	+	335.00	384.32	5.8e-15	1064	! Collagen-like polymer. Recombi

---

A_Geneseq_36:R37741   +   335.00   384.31   5.8e-15   1065   ! Collagen-like polymer DCP3
A_Geneseq_36:R89472   -   335.00   383.95   5.8e-15   1107   ! Collagen/decorin(aa46-93) f
A_Geneseq_36:R89469   -   335.00   383.44   5.9e-15   1169   ! Collagen/BMP-2B fusion prot

seq_name: A_Geneseq_36:W69385

seq_documentation_block:
ID   W69385 standard; Protein; 553 AA.
AC   W69385;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone L1-12 protein.
DE   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy.
KW   Homo sapiens.
OS   WO9837418-A2.
PN   27-AUG-1998.
PD   25-FEB-1998; U03690.
PF   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   N-PSDB; V58586.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Example 1; Page 87-89; 141pp; English.
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence   553 AA;

alignment_scores:
                    Quality: 2861.00   Length: 553
                    Ratio: 5.174   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-030-606-110 x W69385

Align seg 1/1 to: W69385 from: 1 to: 553

284 ATGGTCCAGAGGCTGTGGGTGAGCCGCCTGCTGCGGCACCGGAAAGCCCA 333
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetValGlnArgLeuTrpValSerArgLeuLeuArgHisArgLysAlaGl 17

334 GCTCTTGCCTCGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTTGGCCG 383
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 17 nLeuLeuLeuValAsnLeuLeuThrPheGlyLeuGluValCysLeuAlaA 34

384 CAGGCATCACCTATGTGCCGCCTCTGCTGGAAGTGGGGTAGAGGAG 433
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 34 laGlyIleThrTyrValProProLeuLeuGluValGlyValGluGlu 50

434 AAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTG 483
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LysPheMetThrMetValLeuGlyIleGlyProValLeuGlyLeuValCy 67

484 TGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGCGCTGGGACCGCTATGGCC 533
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 sValProLeuGlySerAlaSerAspHisTrpArgGlyArgTyrArgGlyA 84

534 GCCGGCCGGCCCTTCATCTGGGCACTGCTCCTTGGGCATCCTGCTGAGCCTC 583
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 84 rgArgArgProPheIleTrpAlaLeuSerLeuGlyIleLeuLeuSerLeu 100

Wed Sep 29 14:26:52 1999

```
584  TTTCTCATCCCAAGGGCCGGCTGCTAGCAGGGCTGCTGTGCCGGATCC          633
     |||||||||||||||||||||||||||||||||||||||||||||||||
101  PheLeuIleProArgAlaGlyTrpLeuLeuAlaGlyLeuLeuCysProAspPr     117

634  CAGGCCCCTGGAGCTGGCACTGCTCATCCTGGGCGTGGGCGTGCTGGACT        683
     |||||||||||||||||||||||||||||||||||||||||||||||||
117  oArgProLeuGluLeuAlaLeuLeuIleLeuGlyValGlyLeuLeuAspP       134

684  TCTGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCTGCTCTCTGACCTC        733
     |||||||||||||||||||||||||||||||||||||||||||||||||
134  heCysGlyGlnValCysPheThrProLeuGluAlaLeuLeuSerAspLeu      150

734  TTCCGGACCCGACCACTGCGCCAGGCCTACTCTGTCTATGCCTTCAT         783
     |||||||||||||||||||||||||||||||||||||||||||||
151  PheArgAspProAspHisCysArgGlnAlaTyrSerValTyrAlaPheMe      167

784  GATCAGTCTTGGGGGCTGCCTGGCTACCTCCTGCCATTGACTGGG            833
     ||||||||||||||||||||||||||||||||||||||||||||
167  tIleSerLeuGlyGlyCysLeuGlyTyrLeuLeuProAlaIleAspTrpA       184

834  ACACCAGTGCCCTGGCCCCCTACCTGGCACCCAGGAGAGTGCCTCTTT         883
     ||||||||||||||||||||||||||||||||||||||||||||||||
184  spThrSerAlaLeuAlaProTyrLeuGlyThrGlnGluCysLeuPhe          200

884  GGGCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCACACTGCTGGT       933
     ||||||||||||||||||||||||||||||||||||||||||||||||||
201  GlyLeuLeuThrLeuIlePheLeuThrCysValAlaAlaThrLeuLeuVa       217

934  GGCTGAGGAGGCAGCGCTGGGCCCACGAGCCAGCAGAAGGGCTGTGTCGG       983
     ||||||||||||||||||||||||||||||||||||||||||||||||||
217  lAlaGluGluAlaAlaLeuGlyProThrGluProAlaGluGluCysLeuSerA    234

984  CCCCCTCCTTGTCGCCCCACTCGTGTCCATGCCGGGCCCGTTGGCTTTC        1033
     |||||||||||||||||||||||||||||||||||||||||||||||||
234  laProSerLeuSerProHisCysCysProCysProCysSerArgAlaArgLeuAlaPhe 250

1034 CGGAACCTGGGGCGCCCCTGCTTCCCCGGCTGCTGACCAGCTGCTGCCGCAT     1083
     |||||||||||||||||||||||||||||||||||||||||||||||||||
251  ArgAsnLeuGlyAlaLeuLeuProArgLeuHisGlnLeuCysCysArgMe       267

1084 GCCCCGCCACCCTGCGCCCTCTTCGTGGCTGCGAGCGTGCCAGCTGGATGG     1133
     |||||||||||||||||||||||||||||||||||||||||||||||||||
267  tProArgThrLeuArgLeuPheValAlaGluLeuCysSerTrpMetA         284

1134 CACTCATGAGCCTTCACGCTGTTTTACACGGATTTCGTCGGGCGAGGGGCTG    1183
     ||||||||||||||||||||||||||||||||||||||||||||||||||
284  laLeuMetThrPheThrLeuPheTyrThrAspPheValGlyGluLeu         300

1184 TACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCCAGGCCCGGAGACACTA      1233
     |||||||||||||||||||||||||||||||||||||||||||||||||
301  TyrGlnGlyValProArgAlaGluProGlyThrGlnAlaArgArgHisTy       317

1234 TGATGAAGGCGTTCGGATGGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCA      1283
     |||||||||||||||||||||||||||||||||||||||||||||||||
317  rAspGluGlyValArgMetGlySerLeuGlyLeuPheLeuGlnCysAlaI      334

1284 TCTCCCTGGTCTTCTCCTGTCATGGACCGCTGGTGCAGCCGATTCCGC        1333
     ||||||||||||||||||||||||||||||||||||||||||||||||
334  leSerLeuValPheSerLeuVaLMetAspProGlyLeuValGlnArgPheGly     350

1334 ACTCGAGCAGTCTATTTGGCCCAGTCTGGCCAGCTTTCCCTGGCTGCGCG      1383
     |||||||||||||||||||||||||||||||||||||||||||||||||
351  ThrArgAlaValTyrLeuAlaSerValAlaAlaPheProValAlaAlaGl        367

1384 TGCCACATGCCCTGTCCCCAGTGTGGCCGTGGCCGTGACAGCTTCAGCCGCCC    1433
     ||||||||||||||||||||||||||||||||||||||||||||||||||
367  yAlaThrCysProValProValValAlaValValThrAlaSerAlaAlaL       384

1434 TCACCGGGTTCACCTTCTCAGCCCTGCCTACACCTGGCC               1483
     |||||||||||||||||||||||||||||||||||||||
384  euThrGlyPheThrPheSerAlaLeuGlnIleLeuProTyrThrLeuAla        400

1484 TCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGA      1533
```

```
                                                           |||||||||||||||||||||||||||||||||||||||||
401  SerLeuTyrHisArgGluLysGlnValPheLeuProLysTyrArgGlyAs       417

1534 CACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAG       1583
     ||||||||||||||||||||||||||||||||||||||||||||||||||
417  pThrGlyGlyAlaSerSerGluAspSerLeuMetThrSerPheLeuProG      434

1584 GCCCTAAGCCTGGAGCTCCCTCCTAATGGACACGGTGGTGCTGGAGGC         1633
     ||||||||||||||||||||||||||||||||||||||||||||||||
434  lyProLysProGlyAlaProPheProAsnGlyHisValGlyAlaGlyGly      450

1634 AGTGGCCTGCTCCCACCTCCACCCGGCTCTGCGGGGCCTCTGCCTGTGA      1683
     |||||||||||||||||||||||||||||||||||||||||||||||||
451  SerGlyLeuLeuProProProProAlaLeuCysGlyAlaSerAlaCysAs     467

1684 TGTCTCCGTACGTGTGGTGGGTGAGCCCACCCAGGCCAGGTGGTTC          1733
     ||||||||||||||||||||||||||||||||||||||||||||||
467  pValSerValArgValValGlyGluProThrGluAlaArgValValP        484

1734 CGGGCCGGGCATCTGCCTGGACCTGCCATCCTGGATAGTGCCTTCCTG        1783
     ||||||||||||||||||||||||||||||||||||||||||||||||
484  rGlyArgGlyIleCysLeuAspLeuProIleLeuAspSerAlaPheLeu      500

1784 CTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAG      1833
     ||||||||||||||||||||||||||||||||||||||||||||||||||
501  LeuSerGlnValAlaProSerLeuPheMetGlySerIleValGlnLeuSe      517

1834 CCAGTCTGTCACTGCCTATATGGTGCTGCCGCAGGCTGGGTCTGGTCG       1883
     ||||||||||||||||||||||||||||||||||||||||||||||||
517  rGlnSerValThrAlaTyrMetValSerAlaAlaGlyLeuGlyLeuValA     534

1884 CCATTTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAA       1933
     |||||||||||||||||||||||||||||||||||||||||||||||||
534  laIleTyrPheAlaThrGlnValValPheAspLysSerAspLeuAlaLys       550

1934 TACTCAGCG          1942
     |||||||||
551  TyrSerAla     553
```

seq_name: A_Geneseq_36:W71869

```
seq_documentation_block:
ID   W71869 standard; Protein; 553 AA.
AC   W71869;
DT   06-JAN-1999  (first entry)
DE   Amino acid encoded by prostate tumour clone L1-12.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03492.
PR   09-FEB-1998;  US-020956.
PR   25-FEB-1997;  US-806099.
PR   01-AUG-1997;  US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC; Xu J;
DR   WPI; 98-609886/51.
DR   N-PSDB; V61201.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 1; Page 82-84; 130pp; English.
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence  553 AA;
```

```
alignment_scores:
        Quality:  2861.00          Length:  553
        Ratio:       5.174         Gaps:      0
```

Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-030-606-110 x W71869

Align seg 1/1   to: W71869   from: 1   to: 553

```
284  ATGGTCCAGAGGCGTGGGGTGAGCCGCTGCTGCGGCACCCGAAAGCCCA  333
     ||||||||||||||||||||||||||||||||||||||||||||||||
  1  MetValGlnArgArgLeuTrpIrpValSerArgLeuLeuArgHisArgLysAlaGl   17

334  GCTCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGCCG  383
     ||||||||||||||||||||||||||||||||||||||||||||||||
 17  nLeuLeuValAsnLeuLeuThrPheGlyLeuGluValCysLeuAlaA   34

384  CAGGCATCACCTATGCGCCGCCTCTGCGTGGAAGTGGGGTAGGAGGAG  433
     ||||||||||||||||||||||||||||||||||||||||||||||||
 34  laGlyIleThrTyrValProLeuLeuLeuGluValGlyValGluGlu   50

434  AAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTCTGGGCCTGTCTG  483
     ||||||||||||||||||||||||||||||||||||||||||||||||
 51  LysPheMetThrMetValLeuGlyIleGlyProValLeuGlyLeuValCy   67

484  TGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCC  533
     ||||||||||||||||||||||||||||||||||||||||||||||||
 67  sValProLeuGluGlySerAlaSerAspHisTrpArgGlyArgTyrGlyA   84

534  GCCGCGGCCCTTCATCTGGGCACTCTGCTTGGGCATCCTGCTGAGCCTC  583
     ||||||||||||||||||||||||||||||||||||||||||||||||
 84  rgArgArgProPheIleTrpAlaLeuSerLeuGlyIleLeuLeuSerLeu  100

584  TTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCTGTGCCCGGATCC  633
     ||||||||||||||||||||||||||||||||||||||||||||||||
101  PheLeuIleProArgAlaGlyTrpLeuAlaGlyLeuLeuCysProAspPr  117

634  CAGGCCCCTGAGCTGGCACTGCTCATCCTGGGCGTGGGGCTGCTGGACT  683
     ||||||||||||||||||||||||||||||||||||||||||||||||
117  oArgProLeuGluLeuAlaLeuLeuIleLeuGlyValGlyLeuLeuAspP  134

684  TCTGTGGCCAGGTGTGCTGCTTCACTCCACTGGAGGCCCTGCTCTGACCTC  733
     ||||||||||||||||||||||||||||||||||||||||||||||||
134  heCysGlyGlnValCysPheThrProLeuGluAlaLeuLeuSerAspLeu  150

734  TTCCGGGACCCGACCACTGTGCCCAGGCCTACTCTGTCTATGCCTTCAT  783
     ||||||||||||||||||||||||||||||||||||||||||||||||
151  PheArgAspProAspHisCysCysArgGlnAlaTyrSerValTyrAlaPheMe  167

784  GATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCCATTGACTGGG  833
     ||||||||||||||||||||||||||||||||||||||||||||||||
167  tIleSerLeuGlyGlyCysLeuGlyTyrLeuLeuProAlaIleAspTrpA  184

834  ACACCAGTGCCCTGGCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTT  883
     ||||||||||||||||||||||||||||||||||||||||||||||||
184  spThrSerAlaLeuAlaProThrTrpGlyThrGlnGluGluCysLeuPhe  200

884  GGCCTGCTCACCCTCATCTTCCTCACCTGCTAGCAGCCACACTGCTGGT  933
     ||||||||||||||||||||||||||||||||||||||||||||||||
201  GlyLeuLeuThrLeuIlePheLeuThrCysValAlaAlaThrLeuLeuVa  217

934  GGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCCAGCCAGAGGGCTGTCGG  983
     ||||||||||||||||||||||||||||||||||||||||||||||||
217  lAlaGluGluAlaAlaLeuGlyProThrGluProSerGlnArgLeuSerA  234

984  CCCCCTCCTTGTCGCCACTGTCCATGCCGGGCCCGCTTGGCTTTC  1033
     ||||||||||||||||||||||||||||||||||||||||||||||||
234  laProSerLeuSerProHisCysCysProCysArgAlaArgLeuAlaPhe  250

1034 CGGAACCTGGGGCCCCTGCTTCCCCGGCTCACCCAGCTGTCGTGCCGCAT  1083
     ||||||||||||||||||||||||||||||||||||||||||||||||
251  ArgAsnLeuGlyAlaLeuLeuLeuProArgLeuHisGlnLeuCysCysArgMe  267

1084 GCCCCGGCACCCTGCGCGCCGGCTCTTCCTGGCTGCAGCTGTGCCAGCTGGGATGG  1133
     ||||||||||||||||||||||||||||||||||||||||||||||||
267  tProArgThrLeuArgArgLeuPheValAlaAlaGluLeuCysSerSerTrpMetA  284

1134 CACTCATGACCTTCACGCTGTTTACACGGATTCGTGGGCAGGGGCTG  1183
     ||||||||||||||||||||||||||||||||||||||||||||||||
284  laLeuMetThrPheThrLeuPheThrArgIleArgGlyGlnGlyLeu  300

1184 TACCAGGGCGTGCCCAGCTGAGCCGCCGGCACCCGAGGCCCGAGACACTA  1233
     ||||||||||||||||||||||||||||||||||||||||||||||||
301  TyrGlnGlyValProAlaGluProArgGlyThrArgGlyAlaArgHisTy  317

1234 TGATGAAGGCGTTCGGATGGGCAGCCAGCCTGGGCGTGTTCCTGCAGTGCGCCA  1283
     ||||||||||||||||||||||||||||||||||||||||||||||||
317  rAspGluGlyValArgMetGlySerGlnProGlyValPheLeuGlnCysAlaI  334

1284 TCTCCCTGGTCTTCTCTTCTGGTCATGACCGGCTGGTGCAGCGGATTCGGC  1333
     ||||||||||||||||||||||||||||||||||||||||||||||||
334  leSerLeuValPheSerLeuValMetAspArgLeuValGlnArgPheGly  350

1334 ACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTGTGGCCGTGCCGG  1383
     ||||||||||||||||||||||||||||||||||||||||||||||||
351  ThrArgAlaValTyrLeuAlaSerValAlaAlaPheProValAlaAlaGl  367

1384 TGGCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCC  1433
     ||||||||||||||||||||||||||||||||||||||||||||||||
367  yAlaThrCysLeuSerHisSerValAlaValValThrAlaSerAlaAlaL  384

1434 TCACCGGGTTCACCTTCTCAGCCCTGCCCTACACACTGCCC  1483
     ||||||||||||||||||||||||||||||||||||||||||||||||
384  euThrGlyPheThrPheSerAlaLeuGlnIleLeuProTyrThrLeuAla  400

1484 TCCCTCTACCACCGGGAGAAGCAGTGTTCCTGCCCAAATACCGAGGGGA  1533
     ||||||||||||||||||||||||||||||||||||||||||||||||
401  SerLeuTyrHisArgGluLysGlnValPheLeuProLysTyrArgGlyAs  417

1534 CACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAG  1583
     ||||||||||||||||||||||||||||||||||||||||||||||||
417  pThrGlyGlyAlaSerSerGluAspSerLeuMetThrSerPheLeuProG  434

1584 GCCCTAAGCCTGGAGCTCCCTCCCAATGACACACTGGGTGCTGAGGC  1633
     ||||||||||||||||||||||||||||||||||||||||||||||||
434  lyProLysProGlyAlaProPheProAsnGlyHisValGlyAlaGlyGly  450

1634 AGTGGCCTGCTCCACCCGCCTCGCGGGGCCCTCGCCTGCCTGTGA  1683
     ||||||||||||||||||||||||||||||||||||||||||||||||
451  SerGlyLeuLeuProProProProAlaLeuCysGlyAlaSerAlaCysAs  467

1684 TGTCTCCGTACGTGTGGTGGTGGTGGTGGTGAGCCCACCGAGGCCAGGTGGTTC  1733
     ||||||||||||||||||||||||||||||||||||||||||||||||
467  pValSerValArgValValValValGlyValGlyProThrGluAlaArgValValP  484

1734 CGGGCCGGGGCATTCGCCTGGCCTGGACCTGGCCATCCTGGATAGTGCCTTCCTG  1783
     ||||||||||||||||||||||||||||||||||||||||||||||||
484  roGlyArgGlyIleCysLeuAlaTrpProGlyProAlaIleLeuAspSerAlaPheLeu  500

1784 CTGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCCAGCTCAG  1833
     ||||||||||||||||||||||||||||||||||||||||||||||||
501  LeuSerGlnValAlaProSerLeuPheMetGlySerIleValGlnLeuSe  517

1834 CCAGTCTGCACTGCCTATATGTGTCGCCGGAGGCCTGGGTCTGGTCG  1883
     ||||||||||||||||||||||||||||||||||||||||||||||||
517  rGlnSerValThrAlaTyrMetValSerArgArgArgLeuGlyLeuValA  534

1884 CCATTTACTTTGCTACACCAGGTAGTATTTGACAAGAGCGACTTGGCCAAA  1933
     ||||||||||||||||||||||||||||||||||||||||||||||||
534  laIleTyrPheAlaThrGlnValValPheAspLysSerAspLeuAlaLys  550

1934 TACTCAGCG  1942
     ||||||||
551  TyrSerAla  553
```

seq_name: A_Geneseq_36:W85068

seq_documentation_block:

```
ID   W85068 standard; Protein; 255 AA.
AC   W85068;
DT   12-FEB-1999 (first entry)
DE   Protein encoded by the consensus sequence of the PS108 gene.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; US-850713.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
DR   N-PSDB; V71181.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 17; Pages 99-100; 122pp; English.
CC   The present sequence is encoded by the consensus sequence for a PS108
CC   gene. The sequence is derived from overlapping clones V71166-79. The
CC   clone sequences are PS108 gene-specific. They are used in the method
CC   of the invention. The specification describes a method for detecting the
CC   presence of a target PS108 polynucleotide in a test sample. The method
CC   comprises contacting the test sample with at least 1 PS108-specific
CC   polynucleotide or complement, and detecting the presence of the target
CC   PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence  255 AA;
```

```
alignment_scores:
        Quality:  1287.00        Length:     255
        Ratio:       5.047       Gaps:         0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
us-09-030-606-110  to: W85068  x  W85068

Align seg 1/1  to: W85068  from: 1  to: 255
```

```
1178 GGGCTGTACCAGGGGCGTGCCCAGAGCTGAGCCGGCGACCGGAGGCCCGGAG 1227
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
   1 GlyLeuTyrGlnGlyValProArgAlaGluProGlyThrGluAlaArgAr   17

1228 ACACTATGATGAAGGCGTTCGGATGGGCAGCCTGGGCTGTGTTCCTGCAGT 1277
     |||||||||||||||||||||||||||||||||||||||||||||||||||
  17 gHisTyrAspGluGlyValArgMetGlySerLeuGlyLeuPheLeuGlnC   34

1278 GCGCCATCTCCCTGCTCTTCTCTGGTCATGGACCGGCTGGTGCAGCGA 1327
     |||||||||||||||||||||||||||||||||||||||||||||||
  34 ysAlaIleSerLeuValPheSerValMetAspArgLeuValGlnArg   50

1328 TTCGGCACTCGACCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGTGGC 1377
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  51 PheGlyThrArgAlaValTyrLeuAlaSerValAlaAlaAlaPheProValAl   67

1378 TGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAG 1427
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  67 aAlaGlyAlaThrCysLeuSerHisSerValAlaValAlaValAlaThrAlaSerA   84

1428 CCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCTACACA 1477
     |||||||||||||||||||||||||||||||||||||||||||||||
  84 laAlaLeuThrGlyPheThrPheSerAlaLeuGlnIleLeuProTyrThr  100

1478 CTGGCCTCCCTCACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATACCG 1527
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||
 101 LeuAlaSerLeuTyrHisArgGluLysGlnValPheLeuProLysTyrAr  117

1528 AGGGACACTGGAGGTGCTAGCAGTGAGCAGCCTGATGACCAGCCTTCC 1577
     ||||||||||||||||||||||||||||||||||||||||||||||||
 117 gGlyAspThrGlyGlyAlaSerSerGluAspSerLeuMetThrSerPheL  134

1578 TGCCAGGCCCTAAGCCTGGAGCTCCTTCCCTATGGACACGTGGGTGCT 1627
     ||||||||||||||||||||||||||||||||||||||||||||||||
 134 euProGlyProLysProGlyAlaProPheProAsnGlyHisValGlyAla  150

1628 GGAGGCAGTGGCCTGCTCCCACCTCCACCCGGCCTCTGCGGGGCCTCTGC 1677
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 GlyGlySerGlyLeuLeuProProProProAlaLeuCysGlyAlaSerAl  167

1678 CTGTGATGTCTCCGTACGTGTGTGGTGTGGGTGAGCCACCGAGGCCAGGG 1727
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 aCysAspValSerValArgValValValGlyValGluProThrGluAlaArgV  184

1728 TGGTTCCGGGCCGGGGCATCGCCTGGACCTGCCATCCTGGATAGTGCC 1777
     ||||||||||||||||||||||||||||||||||||||||||||||||
 184 alValProGlyArgGlyIleCysLeuAspLeuAlaIleLeuAspSerAla  200

1778 TTCCTGCGTGTCCCAGGTGGCCCATCCCTGTTATGGGCTCCATTGTCCA 1827
     |||||||||||||||||||||||||||||||||||||||||||||||||
 201 PheLeuLeuSerGlnValAlaProSerLeuPheMetGlySerIleValGl  217

1828 GCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCCAGGCCTGGGTC 1877
     |||||||||||||||||||||||||||||||||||||||||||||||||
 217 nLeuSerGlnSerValThrAlaTyrMetValSerAlaAlaGlyLeuGlyL  234

1878 TGGTCGGCCATTACTTTGCTCACACAGTAGTATTTGACAAGAGCGACTTG 1927
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 euValAlaIleTyrPheAlaThrGlnValValPheAspLysSerAspLeu  250

1928 GCCAAAATACTCAGCG 1942
     ||||||||||||||||
 251 AlaLysTyrSerAla  255
```

```
seq_name: A_Geneseq_36:R80327

seq_documentation_block:
ID   R80327 standard; Protein; 762 AA.
AC   R80327;
DT   18-APR-1996 (first entry)
DE   Protein polymeric adhesion substrate 1-B.
KW   Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;
KW   enzymatic cross-linking; biocompatible material; structural integrity;
KW   medical adhesive; wound closure; tissue repair; transglutaminase.
OS   Synthetic.
PN   WO9523611-A1.
PD   08-SEP-1995.
PF   03-MAR-1995; U02728.
PR   03-MAR-1994; US-205518.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
DR   WPI; 95-320413/41.
PT   Protein polymers comprising repeating units and sequences  - capable
PT   of enzyme-catalysed covalent bond formation useful as a
PT   biocompatible material for wound closure and tissue repair
PS   Example 4; Page 45; 138pp; English.
CC   The amino acid sequence of the protein polymeric adhesion substrate
CC   (PPAS) 1-B. The protein comprises 10 repeats of the PPASl-B monomeric
CC   repeat (R80326) which consists of the CLP 3.7 gene encoded sequence
CC   (R80321) linked to the human fibrin cross-linking substrate peptide
CC   (R80316). The polymers can be used in biological systems
CC   93.2 sequence (R80316). The polymers can be used in biological systems
CC   where in situ formation of a biocompatible material with structural
CC   integrity is required e.g. as medical adhesives and sealants or for wound
CC   closure or tissue repair.
SQ   Sequence  762 AA;

alignment_scores:
```

```
        Quality:   395.00              Length:     854
        Ratio:     1.082               Gaps:        59
Percent Similarity: 42.740    Percent Identity: 30.211

alignment_block:
US-09-030-606-110/rev x R80327   ..

Align seg 1/1  to: R80327   from: 1  to: 762
```

```
2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGG 2298
     ||| ::||:||||||             |||||| ::||::::: |||
  76 GlySerProGly.............AlaProGlyThrProGlyGluGl    87

2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGTAATC 2251
     | :::: ::|||||||||       ::: :|||||:::|
  87 yGlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerP   104

2250 CACCTGCA..........GAGTCCCGCCATTCCAGTGCATGAGGCCCTTCT 2210
     |||:|||              ::||||||  ||| ||||:::|||
 104 rOGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro...  119

2209 GGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTC 2160
     |||||| |||       |||  ::||||||
 120 GlyAlaPro.......GlyThrProGlyPro..GlnGlyLeu...Pr    131

2159 AGGCAGCCCTAGAGACNGTGGGAGAGGAGGACGCCCCAGCCCCCAG 2110
     ||||||||  |||||||| :
 131 oGlySerPro..........GlyAlaProGlyThrProGlyProGlnG   144

2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT......GGCAGCAGAGAG 2066
     ||||||:: ::||| |||                     |||||||||||
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln  159

2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCCAGCCCCGGCAGCCCCA 2016
     ||||:::::|||
 160 GlnHisHisLeuGly.............                       164

2015 TGGGGCTAACAGGAGCGGGGAGCTGGGACCCCAGTGAGCAGCCCTCCAC 1966
     ||||||||:::::|||| |||   |||::||
 165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh  180

1965 CCCAATGTGCTGGAAGTTTCTACCGCTGAGTATTGGCCAAGTCGCTCTT 1916
     |||                            :::
 180 rPro...........................GlyP                183

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGGCCTG 1866
     ||| |||||                    ::||||||
 183 roGlnGlyLeuPro...............GlySerProGlyAla        192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGCTGAGCTGGACAATGGAG 1816
     ||||||                          ||| ||||
 193 ProGlyThrProGlyProGln.......                       199

1815 CCCATAAACAGGGATGGGCCACCTGGGACAGCAGGAAGGCCACTATCCAG 1766
     |||  ||||||::: |||  |||  |||
 200 ..............GlyLeuProGlySerProGlyAla.....ProG    209

1765 GATGGGCCGAGGTCCAGGCAGAGATGCCCGGC.........CCGGAACCA 1728
     || ||||||                          ||| :::
 209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh  225

1727 CCCTGGCCTCCGTGGGCTCACCCACCACCACCAGTACGAGACATCACAG 1678
     | ||| ||| :::|||||                 ||| |||
 225 rProGlyGluGly.....GlnLnHisHisLeuGlyGly.AlaArgGln   239

1677 GCA........GAGGCCCGCAGAGCGCGCGGGTGAGGTGGGAGCAGGCC 1637
     ||| :::||| ::::|||       :::
 240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl  256

1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGAGGAAGGAGCTTCCAGGCTTAG 1587
     |||||||||| :::||
 256 yLeuProGlySerPro........GlyAlaProGlyThr.           266
```

```
1586 GGCCTGCTCAGGAGAGCTGTCATCAGGGTGTCCTGCTCAGCTGCCACCTCCA 1537
     |||||| :::  |||
 267 ..ProGlyProGlnGlyLeu.ProGlySerPro..............    276

1536 GTGTCCCCGGTATTTGGCAGGAACACCTGCTTCCGGGTGGTAGAG 1487
     |||::|||::::|||||| .::||
 277 ..............GlyAlaProGlyThrProGlynG            286

1486 GGAGGCCAGTGTGTAGGGCAGGAGATCTGCAGGGCTGAGAAGGTGAACCCGG 1437
     || ||||             ||||||     ::||||
 286 lyLeuPro...............GlySerProGlyAla.ProGlyThrProGl 298

1436 TGAGGGCGGCCTGAAGCGTCACCAGGCCCACTGTGGGACAGGCCATGTG 1387
     ||||||         :::  |||||
 298 yGluGly.......GlnGlnHisLeuGlyGlyAlaArgGlnAlaGlyA   313

1386 GCACCGGCAG..................CCACAGGGAAAGCTG 1362
     :::||||                    |||||| |||
 313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGG 1312
     |||||||     :::|||||||||||||||
 329 ....................................GlyAlaProGlyTh  337

1311 TCCATGACCAGAGAAGACCAGCAGGAGGATGGCCACTGCAGGAACA..... 1266
     ||| |||||         :::||||||||||
 337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG  354

1265 ...............................GCCCCAGGCTGCCATCCGAA 1245
                                    |||||||  |||::|||
 354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu  370

1244 CGCCTTCATCATAGTGTCTCCGGGCCTCCGGTGCCGGCTCAGCTCTGGGC 1195
     ::::::|||::||||| |||
 371 GlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe  387

1194 A.......CGCCCTGGTACAGCCCCTGCCCACGAAATCCGTGTAAAACAG 1151
     :      ||||||| |||       ||||:::||
 387 rProGlyAlaProGlyThr..ProGlyProGlnGly.............  398

1150 CGTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTGGTGCAGCCGGGAAGC 1101
     :::||||||:::|||       |||::: |||
 399 ...............LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGCGCAGGTGCGGGGCATGGGCAGCACAGCTGGTCAGCCGGGGAAGC 1051
     |||                      ||||:::  |||
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGGCCCAGGTTCC...GGAAAGCCAAGCGGGCCCGGC............ 1014
     ||||||||||:::      ||||| ||||  |||
 427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG  444

1013 ...............................ATGGACAGCAGTGGG     999
                                    ::: |||
 444 lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerProGly 460

 998 GCGACA..AGGAGGGGCCGACAGCCCTTCTGCGTGGTCGGTGGGCCC      952
     |||         ::|||:::|||  |||
 461 AlaProGlyThrProGlyProGlnGlyLeu.........ProGlySerPr 474

 951 AGCGGGTGCTCCTCCAGCCACCAGTGTGGCCTACGCAGGTGAGGAA     902
     |:::: |||  :::|||        ||| |||
 474 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 491

 901 GATGAGGTGAGCAGCCAAAGAGGCACTCCT...............CCT   867
     ||| |||  ::: |||                           |||
 491 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 507

 866 GGGTGCCCAGGTAGGGGCCAGGGCCACTGGTGTCCCAGTCAATGGCAGGC   817
     ||||||||||:::|||       :::  |||||
 508 GlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAl  524

 816 AGGA.......GGTAGCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATA 773
```

```
DE   Collagen-like polymer.
KW   Collagen-like polymer; synthetic polymer; fibre coating;
KW   prosthetic device; catalytic substance.
OS   Synthetic.
PN   US5773249-A.
PD   30-JUN-1998.
PF   02-MAY-1996; 642255.
PR   02-MAY-1996; US-642255.
PR   04-NOV-1986; US-927258.
PR   29-OCT-1987; US-114618.
PR   09-NOV-1988; US-269429.
PR   06-NOV-1990; US-609716.
PR   12-NOV-1991; US-791960.
PR   05-NOV-1992; US-972032.
PR   22-DEC-1995; US-577046.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J, Ferrari FA;
DR   WPI; 98-387004/33.
PT   Recombinant collagen-like polymers - useful for making gels, films,
PT   fibres, etc.
PS   Example 7; Column 51; 93pp; English.
CC   This sequence represents a unnatural collagen-like polymer
CC   of the invention. The products may be used as films, fibres, moulded
CC   objects and admixed with other natural or synthetic polymers or coatings
CC   on fibres, films, labware or other surfaces, e.g. prosthetic devices. The
CC   polymers may be used for binding a wide variety of specific binding
CC   materials, as catalytic substances (where the amino acid sequence may
CC   specifically chelate a wide variety of elements), as purification media,
CC   composites, laminates or adhesives. They may also be combined with
CC   inorganic or organic materials such as carbon fibres, nylon fibres,
CC   nitrocellulose, etc., as flask coatings or in synthetic matrices for the
CC   growth and study of cells, as affinity columns or as supports for
CC   biological materials. The polymers have collagen-like properties, but may
CC   be easily expressed in micro-organisms in high efficiency. The new
CC   sequences can be tailored to give the desired properties.
SQ   Sequence   762 AA;
```

```
alignment_scores:
        Quality:      395.00                 Length:       854
        Ratio:         1.082                  Gaps:          59
   Percent Similarity:  42.740    Percent Identity:  30.211

alignment_block:
US-09-030-606-110/rev x W57668

Align seg 1/1  to: W57668  from: 1  to: 762
```

```
2347  GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGG  2298
      ||:||||||          |||||||   ::||:|||
  76  GlySerProGly........................AlaProGlyThrProGlyGluGl  87

2297  TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATC  2251
      |   :::||||||||          :::  :||||||||:||::|
  87  yGlnGlnHisHisLeuGlyGlyAlaAargGlnAlaGlyAspValGlySerP  104

2250  CACCTGCA........GAGTCCCGCATTCCAGTGCATGGAGCCCTTCT  2210
      ||  ||||          :::||||:||::||          ||||||
 104  roGlyAlaProGlyThrProGlnGlyProGlnGlyLeuProGlySerPro... 119

2209  GGCCTCCCTGTATAAGTCAGACTGAAACCCCTTGGAGGCCTCCAGTC  2160
      |||   |||          |||  ||||||          ::||||||
 120  GlyAlaPro.........GlyThrProGlyPro...GlnGlyLeu...Pr  131

2159  AGGCAGCCCTAGAGACTGGGGAGGAGGAGGGGACGGCCCCAGCCCCCAG  2110
      ||||||||          ::||:||||::||:          |||
 131  oGlySerPro...........GlyAlaProGlyThrProGlyGluGlyGln  144

2109  CTGTGCAGCTACGCACCTCAGCACCACAGGT......GGCAGCAGAGAG  2066
      :::::          :::   ::|||          ||||||
 144  lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln  159

2065  CCACATTACTTTGGCAGCAACAGAAACTGCGGCCCGCAGCCCCA  2016
```

```
                                                          533
524  aGlyAspValGlySerProGlyAlaPro..............

772  GACAGAGTAGGCCTGGGCACAGTGGT...CCGGGT...CCCCGGAAGAGGT  729
     :::      ||||||  :::|||        |||||
534  ....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro  548

728  CAGAGAGCAGGGCCTCCAGTGGAGTGAAGCAACACCTGGCCACAGAAGTCC  679
     ||||||          :::::|||||          :::||
549  GlyThrProGlyProGlnGlyLeuProGlySerProGly.....AlaPr  563

678  AGCAGCCCACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATC  629
     |:::::|||  |||          ::|||||||||          |
563  oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla......P  578

628  CGGGCACAGCAGCCTGCTAGCCACGCCCGCCCCTTGGATGAGAAAGA...  582
     ||||||          :::::  |||||          |||
578  roGlyThrProGlyGluGlyGlnHisHisLeuGlyGlyAlaArgGln  594

581  .........GGCTCAGCAGGATGCCCAAGGACCAGTG...CCCAGATG  547
     |||||          |||::||||  ||          ||||::::
595  AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl  611

546  AAGGGCCGGCGGCGGCCCATAGCTCCACCCCAGTGGTCACTGGCTGAGCC  497
     |||          |||
611  yLeuProGly.................................SerP  616

496  TAGGAGCGGGACACAGACCAGGCCCAGCACTGGACCAATGCCCAGCACCA  447
     ||||||          ||||||
616  roGlyAlaProGlyThrProGlyPro........................  624

446  TGGTCATGAACTTCTCCTCTACCCCACCTTCCAGCAGCAGAGCGGCACA  397
     ::||          |||:::||
625  ...........GlnGlyLeuProGlySerProGlyAla............  633

396  TAGGTGATGCGTCGCGGCCAAACACACCTCCAGCCAAAGGTTAGCAGTT  347
     ||::::|||||          |||  |||
634  ...........ProGlyThrProGlyProGlnGlyLeuProGlyLeuProGlyS  645

346  GACCAGCAAGAGCTGGGCTTTCCGT........................  321
     |||:::  |||
645  erProGlyAlaProGlyThrProGlyGluGlyGlnHisHisLeuGly  661

320  .........GCCGCAGCAGGCGGCTCACCCACCAGCCCTCTGGACCATA  283
     |||::::          ||||||||  |||
662  GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P  678

282  GTGGGCCAGGCGGG...TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAA  236
     |||||||          ||||||          ||||||||
678  roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly  694

235  CTGCTTCGCTCGTCTCGGCTCCAGAAGCTGCGGCCTCTCCTCCTTGCTG  186
     ||||||          ||||||   |||:::::::
695  ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly......  709

185  CCGGCCAACTGCCTAGGAATCAGCCAGGCGCGCCCATTTCTGCCAGCCCTTTG  136
     |||:::::  |||||||          ::|
710  .....ProGlnGlyLeuProGlySerProGlyAla..ProGlyT  722

135  GTGCCGGTCCAGCTTCTCAGCCCATGTCAACACCTGCTGCTGTGGGGCA  86
     |||   |||          |||:::
722  hrProGlyGluGlyGlnHisHisLeuGly...............GlyAla  734

85   CCTCAGTGGGACACGTCTCATCA.........CTCAGATCCTGCCCGA  46
     ||| |||  ||:::|||          :: |||||||||
735  ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg  750
```

```
seq_name: A_Geneseq_36:W57668

seq_documentation_block:
ID   W57668 standard; peptide; 762 AA.
AC   W57668;
DT   27-AUG-1998  (first entry)
```

```
160  |||::::|||
     GlnHisHisLeuGly..............................                164

2015 TGGGGCTAACAGGAGGCGGGGAGCTGGGACCCAGTGAGGCAGGCCTCCAC           1966
     |||||:::::::|||::::|||      |||  |||::|||
165  .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh           180

1965 CCCAAATGCTGCTGGAAGTTTCTACGCTGGCTGAGTATTTGGCCAAGTCGCTCTT      1916
     |||                                            :::
180  rPro.................................................GlyP    183

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGCCAACCCAGGCCTG               1866
     |||   :::||||||
183  roGlnGlyLeuPro................................GlySerProGlyAla 192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGCTGAGCTGGACAATGAG        1816
     ||||||   |||
193  ProGlyThrProGlyGln..................................         199

1815 CCCATAAACAGGGATGGGCACCTGGACAGCAGGAGGCACTATCCAG             1766
     ||||            |||   ||     |||
200  ...............GlyLeuProGlySerProGlyAla......ProG           209

1765 GATGGGCAGGTCCAGGCAGATGCCCCGGC...........CCGGAACCA          1728
     ||   ||||||                                |||  |||
209  lyThrProGlyInGlyLeuProGlySerProGlyAlaPro.GlyTh            225

1727 CCCTGGCTCCGGTGGGCTCACCCACCACCACGTACGGAGACATCACAG           1678
     |||  ::||||||       |||  :::   |||
225  rProGlyGluGly.....GlnInHisLeuGlyGly.AlaArgGln              239

1677 GCA.........GAGCCCCGCAGAGCCGGGTGGGAGGTGGGAGCAGGCC          1637
     |||          ::||||    :::    |||    :::
240  AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl           256

1636 ACTGCCTCCAGCACCGTGTCCATTAGGAGGGAAGGGAGCTCCAGGCTTAG          1587
     |||||||::::|||::::||||||   :::|
256  yLeuProGlySerPro.....................GlyAlaProGlyThr.       266

1586 GGCCTGCGAGGAAGCTGTCATCAGGCTGTCCTCCACTGCTAGCACCTCCA         1537
     ||||||     |||   ::  |||   |||
267  ..ProGlyProGlnGlyLeu.ProGlySerPro.....................      276

1536 GTGTCCCCTCGGTATTTGGGCAGGAACACCTGCTTCTCCCGGTGGTAGAG         1487
     |||::|||:::::|||||||  :::|
277  ..................GlyAlaProGlyThrProGlyProGlnG            286

1486 GGAGGCCAGTGAGTGTGTAGGGCAGGATCTGCAGGAGGTGAAACCCGG          1437
     |||  |||                  :::||||||  |||
286  lyLeuPro..........GlySerProGlyAla.ProGlyThrProGlyProGl       298

1436 TGAGGCGGCGCTGAAGCTGTCACCACGGCCACACTGTGGGACAGGCCATGTG        1387
     ||||||      :::  ::  |||    |||
298  yGluGly.....GlnInHisLeuGlyGlyAlaArgGlnAlaGlyAlyA            313

1386 GCACCGGCAG.....................CCACAGGGAAAGCTG            1362
     ::||||            |||||||||     |||
313  spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu          328

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGG         1312
     |||||:::::::||||||                  |||  |||:::||
329  ProGlySerPro...............................GlyAlaProGlyTh     337

1311 TCCATGACCAGAGAAGAAGAGAGGGAGGAGATGGCCACTGCCAGGAACA          1266
     |||  |||     :::     ||||||
337  rProGlyInGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlyThrProG    354

1265 ..................GCCCCAGGCTGCCCATCCGAA                 1245
     |||||||||     :::||
354  lyProGlnGlyLeuProGlyAlaProGlyThrProGlyGlu                  370

1244 CGCCTTCATCATAGTGTCTCCGGGCCTTCGGTGCCCGGCCTCAGCTCTGGGC      1195
     :::|||  :::::::|||      ::
```

```
371  GlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe         387

1194 A.............CGCCCTTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG  1151
     |||||||||     |||::|||      |||  |||::|||
387  rProGlyAlaProGlyThr..ProGlyProGlnGly..................      398

1150 CGTGAAGGTCATGAGTGCCATCACCAGCTGCACCAGCTCAGCCACGAAGAGCC      1101
     :::|||||:|||
399  .........LeuProGlySerProGlyAlaProGlyThrProGly              410

1100 GGGCCAGGTGCGGGGCATGGGCGAGCACAGCTGGTGCAGCCCGGGGAAGC        1051
     |||    :::       ||::|||   |||
411  ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl         427

1050 AGGGCGCCAGGTTCC...GGAAAGCCAAGCGGGCCCGGC..............      1014
     |||  |||::|||      |||  |||::|||
427  nGlyLeuProGlyProGlyAlaProGlyThrProGlyGluGlyGlnInG         444

1013 ...............................ATGGACAGCAGTGGG           999
     :::|||
444  lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerProGly         460

998  GCGACA...AGGAGGGGCCGACAGCCCTTCGCTGGCTCGGTGGGGCCC          952
     |||      ||||||      :::|||
461  AlaProGlyThrProGlyProGlnGlyLeu..........ProGlySerPr        474

951  AGGGCTGCCTCCTCAGCCACCACCAGCAGTGGGCTGCTACGCAGGTGAGGAA      902
     |:::     ::||||||       |||  :::   |||
474  oGlyAlaProGlyThrProGlnGlyLeuProGlnGlyLeuProGlySerProGlyA   491

901  GATGAGGGTGAGCAGCAGCCAAAGAGGCACTCCT...............CCT      867
     |||   |||   :::||||         ||||||            |||
491  laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro         507

866  GGGTGCCCAGGTAGGGGCCAGGGCACTGGTGTCCAGTCAATGGCAGGC          817
     ||||||   |||                ||   |||
508  GlyThrProGlyGlnGlyGlnInHisLeuGlyGlyAlaArgGlnAl           524

816  AGGA..........GGTAGCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATA    773
     ||||          :::|||
524  aGlyAspValGlySerProGlyAlaPro...............................  533

772  GACAGAGTAGGCCTGGCCGACAGTGGT...CCGGGT...CCCGGAAGAGGT       729
     ||||||      :::||||||      |||||||
534  .....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro         548

728  CAGAGAGCAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCC        679
     |||||          :::||||||||     :::||
549  GlyThrProGlyProGlnGlyLeuProGlySerProGly.....AlaPr          563

678  AGCAGCCCCACGCCCAGGATGAGCAGTGCCAGTCCAGGGGGCCTCAGGGATC      629
     :::||||||||     |||   ::||||||||
563  oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla......P        578

628  CGGGCACAGCAGCCCTGCTAGCCAGCCGCCCTTGGGATGAGAAAGA...        582
     ||||||||     |||   :::   |||   |||
578  roGlyThrProGlyGluGlyGlnInHisLeuGlyGlnAlaArgGln             594

581  .........GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATG          547
     |||||||      |||      |||  |||:::
595  AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl          611

546  AAGGGCCCGCGCGGCCCATAGCGTCCACGCCCAGTGGTCACTGGCTGACCC       497
     |||      |||
611  yLeuProGly...............................SerP            616

496  TAGGAGCGGGACACAGGCCACAGGCCCAGCCACTGGACCAATGCCCAGCACCA      447
     |||      ||||||||
616  roGlyAlaProGlyThrProGlyPro............................      624

446  TGGTCATGAACTTCTCCTCTACCCCCATTCCAGCAGCAGAGGGCGGCACA         397
     |||      ||||||:::|||
625  ...............GlnGlyLeuProGlySerProGlyAla............      633
```

```
alignment_scores:
                   Quality:  395.00                    Length:    854
                     Ratio:    1.082                     Gaps:      59
        Percent Similarity:   42.740    Percent Identity:  30.211

alignment_block:
US-09-030-606-110/rev x W49715

Align seg 1/1  to: W49715  'from: 1  to: 762

2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGG 2298
      |||::::||||||          |||||||     ::||  ::||||::: ||
  76 GlySerProGly..........AlaProGlyThrProGlyGluGl       87

2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATC 2251
      |   ::: |||||||||||     :: |:|||||||||||:::||:::::|
  87 yGlnGlnHisHisLeuGlyGlyAlaGlnAlaGlyAlaAspValGlySerP 104

2250 CACCTGCA........GAGTCCCGCATTCCAGTGCATGGAGCCCTTCT 2210
      ||  |||          :::|||||||:::::    |||||||||||||
 104 roGlyAlaProGlyThrProGlnGlyProGlyLeuProGlySerPro... 119

2209 GGCCTCCCGTGTATAAGTCCAGACTGAAACCCCCTTGGAAGGCCTCCAGTC 2160
      ||  |||                 |||:::||||||      ||  ::|||||||
 120 GlyAlaPro........GlyThrProGlyPro...GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGGAGGAGAGGAGGACGCCCCAGCCCCCCCAG 2110
      ||||||||||                ||||||||  |||
 131 oGlySerPro..........GlyAlaProGlyThrProGlyProProGlnG 144

2109 CTGTGCAGCTACGGCACCTCCACCTCAGCAGCACAGGT.....GGCAGCAGAGAG 2066
      ::::      :::   ::||| |||   :||||||||||||     ||||||||||
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159

2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCAGCCCCCA 2016
      ||||::::::|||
 160 GlnHisHisLeuGly..................................... 164

2015 TGGGGCTAACAGAGACGCGGGAGCTGGGACCCAGTGAGCCTGAGCGCCTCCAC 1966
      |||||::::::||:::::||||     ||  ||||||||||:|||:|
 165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180

1965 CCCAATGTGCTGGAAGTTTCTACGCTGGAGTATTGGCCAAGTCGCTCTT 1916
      ||||||                               :::
 180 rPro..................................GlyP 183

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGGACCCAGGCCTG 1866
      |||||||                          ::||||||
 183 roGlnGlyLeuPro...................GlySerProGlyAla 192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816
      ||||||
 193 ProGlyThrProGlyGln................................. 199

1815 CCCATAAACAGGATGGGCCACCTGGACAGGAGGCACTATCCAG.....ProG 1766
      |||  ||||||:::   |||           ||||
 200 ..................GlyLeuProGlySerProGlyAla......ProG 209

1765 GATGGCGAGTCCAGGCAGATGCCCCGGC...........CCGGAACCA 1728
      ||  ||||||||||||||||:::          |||  :::
 209 lyThrProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225

1727 CCCTGGCCTCGGTGGCTCACCCACCACACGGAGAGATCACAG 1678
      |||||||||     |||  |||::::
 225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlaArgGln 239

1677 GCA........GAGGCCCCGCAGAGCGCGGGGTGGAGGTGGGGAGCAGGCC 1637
      |||          ::|||    :::    |||   |||   ||||
 240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlnGl 256

1636 ACTGGCTCCAGCACCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTAG 1587
      ||||||||||
 256 yLeuProGlySerPro..................GlyAlaProGlyThr. 266

 396 TAGGTGATGATGCTCGGCGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTT 347
      |||::||||||          |||     |||   |||
 634 ..............ProGlyThrProGlyProGlnGlyLeuProGlyS 645

 346 GACCAGCAAGAGCTGGGCTTCCGGT................... 321
      |||::|        |||
 645 erProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGly 661

 320 ..............GCCGCAGCAGGCGGCTCACCCACAGCCTCTGACCATA 283
      ||||::::        ||||||||     |||      ||
 662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678

 282 GTGGGCCAGGCGGG...TAGGGCTCAGGGCGGCCGTTCAGGCACTCCAGAA 236
      ||||||        ||       |||||||||     ||||||||||
 678 roGlyProGlnGlyLeuProGlyAlaProGlyThrProGly 694

 235 CTGCTTCGTCTCGGCTCGCTCGCTCCAGAAAGCTGCGGCCTTCCTCCTGCTG 186
      ||  ::::||| ||| :::||||||
 695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly...... 709

 185 CCGGCAACTGCCTAGGAATCAGCCAGGCGGCCCATTTCTGCCAGCCCTTTG 136
      |||::::     ||||||:: ::   |||        :::   ||
 710 ..........ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722

 135 GTGCCGGTCCAGCTTCTCAGCCCCATGCTCAACACCTGCTGCTGTGGGGCA 86
      ||      :::|||||| ||      ||||||||||
 722 hrProGlyGluGlyGlnGlnHisHisLeuGly.......... 734

  85 CCTCAGTGGGGACACGTCTCCATCA........CTCAGAATCCTGGCCGA 46
      ||  ||||||||||||| :::      |||||||||||
 735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
```

seq_name: A_Geneseq_36:W49715

```
seq_documentation_block:
ID   W49715 standard; Protein; 762 AA.
AC   W49715;
DT   12-OCT-1998 (first entry)
DE   Protein polymer adhesive substrate PPAS1-B.
KW   Protein polymer adhesive substrate; PPAS1-B; sealant;
KW   wound healing; CLP 3.7; collagen; human.
OS   Synthetic.
OS   Homo sapiens.
PN   US5773577-A.
PD   30-JUN-1998.
PF   03-MAR-1994; 205518.
PR   02-MAR-1995; US-397633.
PR   03-MAR-1994; US-205518.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
DR   WPI; 98-387091/33.
PT   New recombinant protein polymers - containing naturally occurring
PT   repetitive units for crosslinking by enzymes, useful as medical
PT   adhesives and sealants, depots and matrices
PS   Example 4; Column 27; 70pp; English.
CC   Protein polymer adhesive substrate PPAS1-B comprises 10 repeats
CC   of a monomer (see W49714) consisting of human fibrin gamma chain
CC   peptide 93.2 (see W49708) inserted into a structural backbone of
CC   collagen repeat motifs (see W49711). The 68 kDa PPAS1-B polymer
CC   has been expressed in Escherichia coli HB101 transformants using
CC   plasmid pPT0324. Claimed recombinant protein polymers are capable
CC   of covalent crosslinking by enzymatic reaction to form products
CC   which set quickly and have good adhesive properties and high
CC   strength. CLP (collagen like protein) polymers are extremely soluble
CC   in water, allowing protein solutions of over 10 wt.% to be formed
CC   while maintaining good flow properties. CLP polymers have good
CC   adhesion to hydrophilic surfaces and therefore may adhere well to
CC   tissue. The fibrin peptide 93.2 inserts of PPAS1-B provide
CC   glutamine donor sites for Factor XIIIa mediated cross-linking.
SQ   Sequence 762 AA;
```

```
1586 GGCCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCA 1537
 267 ..ProGlyGlnGlyLeu.ProGlySerPro................ 276

1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCCCCGGTGGTAGAG 1487
 277 .........GlyAlaProGlyThrProGlyProGlnG 286

1486 GGAGGCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAGGTGAACCCGG 1437
 286 lyLeuPro........GlySerProGlyAla.ProGlyThrProGl 298

1436 TGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGTGGGACAGGCATGTG 1387
 298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyA 313

1386 GCACCGGCAG...........................CCACAGGGAAAGCTG 1362
 313 spValGlySerProGlyAlaProGlyThrProGlnGly...Leu 328

1361 CCACACTGGCCAAAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGG 1312
 329 ProGlySerPro.........................GlyAlaProGlyTh 337

1311 TCCATGACCAGAGAGAGACCAGGAGATGGCGCACTGCAGGAACA.... 1266
 337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProg 354

1265 ...............................GCCCCAGGCTGCCCATCCGAA 1245
 354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTGTCTCCGGGGTCTCCGGTGCCCGGTCAGCTCTGGGC 1195
 371 GlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe 387

1194 A......CGCCCTGGTACAGCCCCTCGCCCCACGAAATCCGTGTAAAACAG 1151
 387 ;..........ProGlyProGlnGly.......ProGlyProGlnGly.. 398

1150 CGTGAAGGTCATGAGGTGCCATCCAGGTGCACAGCTCAGCCACGAAGAGCC 1101
 399 ..........LeuProGlySerProGlyAlaProGlyThrProGlyGlu 410

1100 GGGCAGGGTGCGGGGCATGCGGGCAGCAGCTGGTGCAGCCGGGGGAAGC 1051
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGCGCCCAGGTCC...GGAAAGCCAAGCCGGCCCGGC............. 1014
 427 nGlyAlaProGlyThrProGlyLeuProGlyProGlyGluGlyAspValG 444

1013 ....................................ATGGACAGCAGTGGG 999
 444 lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerProGly 460

 998 GCGACA...AGGAGGGGGCCGACAGCCCTTCTGCTGGTCGGTGGGGCCC 952
 461 AlaProGlyThrProGlyGlnGlnHisHisLeu........ProGlySerPr 474

 951 AGCGCTGCCTCCTCAGCCACCAGTGTGGCTACGCAGGTGAGGAA 902
 474 oGlyAlaProGlyThrProGlyLeuProGlnGlyLeuProGlySerProGlyA 491

 901 GATGAGGGTGAGCAGCCCAAAGAGGCACTCC............CCT 867
 491 laProGlyThrProGlyGlnGlyLeuProGlySerProGlyAlaPro 507

 866 GGGTGCCCAGTAGGGGGCCACTGGTCTCCCAGTCAATGGCCAGGC 817
 508 GlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAl 524
```

```
 816 AGGA......GGTAGCCCAGGCAGCCCCAAGACTGATCATGAAGGCATA 773
 524 aGlyAspValGlySerProGlyAlaPro................::::::: 533

 772 GACAGAGTAGGCCTGGGCGACAGTGGT...CCGGGT...CCCGGAAGAGGT 729
 534 ....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 548

 728 CAGAGAGCAGGGCCTCCAGTGAGTGAAGCCACACCTGGCCACAGAAGTCC 679
 549 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.....AlaPr 563

 678 AGCAGCCCCACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCTGGGATC 629
 563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla......P 578

 628 CGGGCACAGCAGCCCTGCTAGCCAGCCCGCCCCTGGGATGAGAAAGA 582
 578 roGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGln 594

 581 ...........GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATG 547
 595 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlnGl 611

 546 AAGGGCCCGGGCGGGGTCCACGCCAGTGGTCACTGGCTGAGCC 497
 611 yLeuProGly................................SerP 616

 496 TAGGAGCGGGACAGACACCAGGCCCAGCACTGGACCAATGCCCAGCACCA 447
 616 roGlyAlaProGlyThrProGlyGlu.......... 624

 446 TGGTCATGAACTTCTCCTCTACCCCACTTCCAGCAGCAGAGGCGGCACA 397
 625 ................GlnGlnHisHisLeuGlyGlyAlaArgGlnAla... 633

 396 TAGGTGATGCCTGCGGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTT 347
 634 .........ProGlyThrProGlyProGlnGlyLeuProGlyLys 645

 346 GACCAAGGAGAGCTGGGGCTTTCCGGT................ 321
 645 erProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGly 661

 320 ...............GCCGAGCAGGCGGCTCACCCACCACAGCCTCTGGACCATA 283
 662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678

 282 GTGGGCCAGGCGGG...TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAA 236
 678 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 694

 235 CTGCTTCCTCTCCTGCTCCTGCCTCAGAAGCTGCGGCCCTCTCCTCCTTGCTG 186
 695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly...... 709

 185 CCGCCAACTGCCTAGGAATCAGCCAGGCGGCCCCATTCTGCCAGCCCCTTG 136
 710 ........ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722

 135 GTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCA 86
 722 hrProGlyGluGlyGlnGlnHisHisLeuGly..............GlyAla 734

  85 CCTCAGTGGGACACGTCTCATCA.......CTCAGATCCTGGCCGA 46
 735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
```

seq_name: A_Geneseq_36:R80324

seq_documentation_block:
ID   R80324 standard; Protein; 762 AA.
AC   R80324;

```
DT   18-APR-1996  (first entry)
DE   Protein polymeric adhesion substrate 1-A.
KW   Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;
KW   enzymatic cross-linking; biocompatible material; structural integrity;
KW   medical adhesive; wound closure; tissue repair; transglutaminase.
OS   Synthetic.
PN   WO9523611-A1.
PD   08-SEP-1995.
PF   03-MAR-1995; U02728.
PR   03-MAR-1994; US-205518.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
DR   WPI; 95-320413/41.
PT   Protein polymers comprising repeating units and sequences  - capable
PT   of enzyme-catalysed covalent bond formation useful as a
PT   biocompatible material for wound closure and tissue repair
PS   Example 4; Page 43; 138pp; English.
CC   The amino acid sequence of the protein polymeric adhesion substrate
CC   (PPAS) 1-A. The protein comprises 10 repeats of the PPASI-A monomeric
CC   repeat (R80323) which consists of the CLP 3.7 gene sequence linked to the
CC   human fibrin cross-linking substrate sequence (see Q98704) corresponding
CC   to the wild type 17 C-terminal amino acids from human fibrin. The
CC   polymers can be used in biological systems where in situ formation of a
CC   biocompatible material with structural integrity is required e.g. as
CC   medical adhesives and sealants or for wound closure or tissue repair.
SQ   Sequence  762 AA;

alignment_scores:
     Quality:    393.00             Length:     852
     Ratio:        1.059            Gaps:        59
Percent Similarity:  43.545         Percent Identity:  30.516

alignment_block:
US-09-030-606-110/rev x R80324  ..

Align seg 1/1  to: R80324  from:  1  to: 762
```

```
1100 GGCGCAGGGTCGCGGGGCATGCGGCAGCAGCTGGTCGTGCGGGGAAGC 1051
              :::        ::::::       |||
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlyProGl 427

1050 AGGGGCGCCAGGTTCC...GGAAAGCCAAGCGGGCCCGGCAT.GGACAGC 1005
     ||||||  |||||||||  ||  ||||:::  |||||| ||||||
 427 nGlyLeuProGlySerProGlyAlaProGlyThrProGluGluGlnG 444

1004 AG.........TGGGGGCGACAAGGAGGGGGCGACAGCCCTTCTGCTGGC 964
     ||        ||  |||||||||||||||| ||||| |||  :::|||
 444 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 460

 963 TCGGTGGGG..............................CCCAGCGCTGC 944
     :::                                     |||:::||||
 461 AlaProGlyThrProGlyProGlyLeuProGlySerProGlyAla.P 477

 943 CTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGG 894
     ||  |||      |||                       |||
 477 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 493

 893 TGAGCAGGCCAAAGAGGCACTCCT.............CCTGGGTGCCC 859
     ||  :::|||  |||||| ||||||             ||||||||||
 494 ThrProGlyProGlnGlyLeuProGlySerProGlyProGlyAlaProGlyThrPr 510

 858 AGGTAGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA.... 813
     |||:::||||  |||      :::           |||||||||
 510 oGlyGluGlyGlnGlnHisLeuGlyAlaAlaLysGlnAlaGlyAlaLysGlnAlaGlyAspV 527

 812 ..GGTAGCCCAGGCAGCGCCCCCAAGACTGATCATGAAGGCATAGACAGT 765
     ||||||||||||||:::|||                        :::
 527 alGlySerProGlyAlaPro............................Gly 534

 764 AGGCCTGGCGACAGTGT...CCGGGT...CCCGGAAGAGGTCAGAGAGC 721
     ||||||  ||||||  |||||| ||||||  |||||||
 535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551

 720 AGGGGCTTCCAGTGGAGTGAAGCACACTGGCCACAGAAGTCCAGCAGCCC 671
     ||||||      :::::|||||    :::|||::::::|
 551 oGlyProGlnGlyLeuProGlySerProGly......AlaProGlyThrProP 566

 670 CACGCCCAGGATGAGCAGTGCCAGTCCAGGGGATCCGGGCACA 621
     ||  |||    :::|||||||||        |||||||||||
 566 roGlyProGlnGlyLeuProGlySerProGlyAla......ProGlyThr 580

 620 GCAGCCCTGCTAGCCAGCCCTGGATGAGAAGA............ 582
     :::  ||||||  :::||        |||||||
 581 ProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaAlaLysGlnAlaGlyAs 597

 581 ...GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATGAAGGGCCG 539
     ||||||  |||::::|||  |||      |||:::
 597 pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 614

 538 GCGGGCCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGGCG 489
     ||     |||||||
 614 ly......................................SerProGlyAla 618

 488 GGACACCAGACCAGGCCCAGCACTGACCAATGCCCAGCACCATGGTCATG 439
              ||||||||||
 619 ProGlyThrProGlyPro................................ 624

 438 AACTTCTCCTCACCCCCACTTCCAGCAGCAGAGGCGGCACATAGGTGAT 389
     :::       |||||||   ||::||
 625 ....GlnGlyLeuProGlySerProGlyAla.................... 633

 388 GCCTGCGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339
     ||||::::||||   |||  ||  ||:::||
 634 .......ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647

 338 AGAGCTGGGCTTCCGGT............................. 321
     |||               |||
 648 AlaProGlyThrProGlyGluGlyGlnGlnHisLeuGlyGlnHisHisLeuGlyGlyAlaAlaLy 664

 320 ....GCCGCAGCAGGGCGGCTCACCCACAGCCCTCTGGACCATAGTGGGCCA 275
              :::|||||||||     :::    :::|||:::|||
 687 yGlnGlnHisHisLeuGlyGlyAlaAlaLysGlnAlaGlyAspValGlySerP 104
```

```
 664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680
     |||:::          |||||||    |||    ||||||
 274 GGCGGG...TAGGGCTCAGGGGGCCGTTCAGGCGACTCCAGAACTGCTTCG 228

 681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl 697
           ||| ||||||    ||  |||||||
 227 TCTCGGCTCGTGCTCCAGAAGCTCGCGGCCTCTCCCTTCCTTGCTGCCGCCAAC 178

 697 yLeuProGlySerProGlyAlaProGlyThrProGly............. 709
     |||  :::::|||  |||  :::::|||
 177 TGCCTAGGAATCAGCCAGGCGCCCATTCTGCCAGCCCTTGTGTGCCGGT 128

 710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724
          |||  :::::|||    :::        |||||||
 127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTG 78

 725 GluGlyGlnGlnHisHisLeuGly...........GlyAlaLysGlnAl 737
     :::::|||                       ||||||
  77 GGGACACGTTCCATCA.........CTCAGATCCTGGCCGA 46

 737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
     |||  |||:::||||||||||||
  37 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
```

seq_name: A_Geneseq_36:W53535

seq_documentation_block:
```
ID   W53535 standard; Protein; 762 AA.
AC   W53535;
DT   10-AUG-1998  (first entry)
DE   Amino acid sequence of the plasmid encoding fibrin gamma PPAS1-A.
KW   fibrin gamma PPAS1-A; peptide repeat unit; DNA repeat unit;
KW   high molecular weight polymer; synthetic silk; silk worm.
OS   Synthetic.
PN   WO9810063-A1.
PD   12-MAR-1998.
PF   23-SEP-1996; U15306.
PR   03-SEP-1996; US-707237.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J, Crissman JW, Dorman MA, Ferrari FA;
DR   WPI; 98-193613/17.
PT   Preparation of synthetic repetitive DNA - useful for construction of
PT   large protein polymers having repeating units, used in structural
PT   material, e.g. synthetic silk
PS   Example 6; Pages 83; 127pp; English.
CC   This is the amino acid sequence of the fibrin gamma PPAS1-A encoded
CC   on plasmid pPT0321, used in the method of invention, which involves
CC   the preparation of synthetic DNA sequence having repeating units from
CC   about 3-15 codons and encoding a protein of at least about 30 kDa.
CC   The method is useful for the production of high molecular weight
CC   polymers (e.g. synthetic silk), either nucleic acids or peptides that
CC   are the expression products of the nucleic acids and particularly high
CC   molecular weight peptides containing repeating units which are useful
CC   structural materials.
SQ   Sequence  762 AA;
```

alignment_scores:
```
    Quality:  393.00           Length:     852
    Ratio:      1.059           Gaps:        59
Percent Similarity: 43.545   Percent Identity: 30.516
```

alignment_block:
US-09-030-606-110/rev x W53535

Align seg 1/1 to: W53535 from: 1 to: 762

```
2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGG 2298
     ||||:::||:::    |||    ||::::  ::||
2347 GlySerProGly.............AlaProGlyThrProGlyGluG 87

2297 TGTGTCT...CAACTAGGAGGCTAGCGTGTTAACCCTGCAGCCTGGGTAATC 2251
     |        ::||||||||       :::  ||||||||||:::::|
  87 yGlnGlnHisHisLeuGlyAlaAlaLysGlnAlaGlyAspValGlySerP 104
```

```
                                                 CCACAGGGAAAGCTG 1362
                                                 |||||||
1386 GCACCTGCAG.......                           ::::|||
313       spValGlySerProGlyAlaProGlyThrProGlnGly...Leu 328

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCAATCGTCGACCAGCCGG 1312
                                                 :::||||||||:::
329       ProGlySerPro.....                     .GlyAlaProGlyTh 337

1311 TCCATGACCAGAGAGAAGACCAGGGAGGATGGCCACTGCAGGAACA.... 1266
     ||| |||:::  ||||||                ::: |||||
337       rProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354

1265 .........                   GCCCCAGGCTGCCCATCCGAA 1245
                                |||:::::|| |||::|||
354       lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTGTCTCCGGCCCTCGGTCCCGGCTCAGCTCTGGGC 1195
     :::|||              ::: ::
371       GlyGlnGlyHisHisLeuGlyGlyAlaAlaLysGlnAlaGlyAspValGlySe 387

1194 A......CGCCCTGGTACAGCCCCTGCCCCACGAAATCCGTGTAAAACAG 1151
     :          ||||||| |||              ||:::::
387       rProGlyAlaProGlyThr..ProGlyProGlnGly.......         398

1150 CGTGAAGGTCATGAGCTGCCATCCAGCTGCACAGTCCCAGCCACGAAGAGCC 1101
     :::|||              ||| :::              ::::|||
399       ........LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGGCCAGGGTGCGGGGCATGCGGCAGCAGCTGGTGCAGCCCGGGGAAGC 1051
     |||              :::::: ||| :::::: |||||
411       ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGGCGCCCAGGTTCC...GGAAAGCCAAGCGGCCCGGCCAT.GGACAGC 1005
     ||| :::::: |||  :::::: |||::|| |||||
427       nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444

1004 AG.......TGGGGCGACAAGGAGGGGGCCCGACAGCCCTTCTGCTGGC 964
     ||          |||:::::::::|||              |||
444       lnHisHisLeuGlyAlaLysGlnAlaGlyAspValGlySerProGly 460

963  TCGGTGGGG.......              CCCAGCCGCTGC 944
     :::| |||              |||::::||:.P
461       AlaProGlyThrProGlnGlyLeuProGlySerProGlyAla.P 477

943  CTCCTCAGCCACCAGCAGTGTGGCTGCTACGGCAGGTGAGGAAGATGAGGG 894
     |||  :::::::|||||              |||
477       rProGlyThrProGlnGlyLeuProGlySerProGlyAlaProGlyGlnG 493

893  TGAGCAGGCCAAAGAGGCACTCCT.........              CCTGGTGCCC 859
     ||| :::::| |||              |||||||
494       ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510

858  AGGTAGGGGGCCCAGGGCACTGGTCTCCCAGTCAATGGCAGGCAGGA.... 813
     ||| ::: |||              ||||||||
510       oGlyGluGlyGlnGlyHisHisLeuGlyGlyAlaAlaLysGlyAspV 527

812  ..GGTAGCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGT 765
     :|||||              :::||||| |||
527       alGlySerProGlyAlaPro.........              Gly 534

764  AGGGCTGGCGACAGTGGT...CCGGGT...CCCGGAAGAGGTCAGAGAGC 721
     ||||||  :::||| |||||              |||||
535       ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551

720  AGGGGCTTCCAGTGAGTGAAGCACACTGGCCACAGAAGTCCAGCAGCCC 671
     |||||              :::|||||||              ||||
551       oGlyProGlnGlyLeuProGlySerProGlyAla.....AlaProGlyThrP 566

670  CACGGCCCAGGATGAGCAGTGCCAGCTCCAGGGGGATCCGGGCCACA 621
     ||| |||              :::|||||||              |||||
566       roGlyProGlnGlyLeuProGlySerProGlyAla......ProGlyThr 580

620  GCAGCCCTGCTAGCCGACCGCCCCTTGGGATGAGAAAGA......... 582
```

```
2250 CACCTGCA........GAGTCCCGCCATTCCAGTGCATGCGAGCCCTTCT 2210
     ||  |||              ::::||:::::
104       roGlyAlaProGlyThrProGlnGlyLeuProGlySerPro... 119

2209 GGCTCCCGTATAAGTCCAGACTGAAACCCCTTGGAAGGCTCCAGTC 2160
     ||| |||              ::::||||||
120       GlyAlaPro.....              .GlyThrProGlyPro..GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGAGGAGAGGAGGACGCCCCAGCCCCCAG 2110
     ||||||||              |||||||| |||   :
131       oGlySerPro.........              GlyAlaProGlyThrProGlyProGlnG 144

2109 CTGTGCAGCTACCGCACCTCAGCACCAGGGT.....GGCAGCAGAGAG 2066
     :::::              ::: ||:::|||              :::
144       lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGlnGln 159

2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCGACCCGGCCAGCCCCA 2016
     |||::::|||
160       GlnHisLeuGlyHis.........              164

2015 TGGGCTAACAGGAGCGGGGAGGTGGGACCCAGTGAGGCAGGCCCTCCAC 1966
     |||||||::::::::|| ||| |||:::||
165       .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180

1965 CCCAATGTGCTGGAAGTTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTT 1916
     ||||              :::
180       rPro.........              .GlyP 183

1915 GTCAAATAGTACCTGTGTAGCAAAGTAAATGGGCACCAGACCCAGGCCTG 1866
     ::|||||||              |||
183       roGlnGlyLeuPro.........              GlySerProGlyAla 192

1865 CGGCAGCACCACCATATAGGACAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816
     ||||||              |||
193       ProGlyThrProGlyGln.........              199

1815 CCCATAAAACAGGGGATGGGGACCACCTGGGACAGCAGGAAGGCACTATCCAG 1766
     ||| |||||||:::              ||| ||||
200       ........              GlyLeuProGlySerProGlyAla......ProG 209

1765 GATGGCGAGGTCAGGCAGCAGATGCCCCGGC..........CCGGAACCA 1728
     ||||||              ::::: |||||||
209       lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225

1727 CCCTGGCCTCGGTGGGCTCACCCCACCACCACGTACGGAGACATCACAG 1678
     ||||||              ::|||||| :::
225       rProGlyGluGlyGln.........              GlnHisHisLeuGlyGly.AlaLysGln 239

1677 GCA........GAGGCCCCAGAGCGCGGGTGGGAGGTGGGAGCAGGCC 1637
     |||              ::::|| :::
240       AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 256

1636 ACTGCCTCCAGCACCCACCGTGTCCATTAGGGAAGGGAGGTTCCAGGCTTAG 1587
     ||||||||||
256       yLeuProGlySerPro.........              GlyAlaProGlyThr. 266

1586 GGCCCTGGCAGGAAGCTGGTCGTCATCAGGCTGTCCTCACTGCTAGCACCTCCA 1537
     ||||||  :::  :::|| |||
267       ..ProGlyProGlnGlyLeu.ProGlySerPro......... 276

1536 GTGTCCCCTCCGGTATTTGGGCAGGAACACCTGCTTCTCCCCGGTGGTAGAG 1487
     |||::::::||:::::||||||              ::|
277       .........              GlyAlaProGlyThrProGlyProGlnG 286

1486 GGAGGCCAGTGTGTAGGGTCGAGGGCTGCAGGGCTGAGAAGGTGAACCCGG 1437
     ||||||              :::::||||||              |||||
286       lyLeuPro.........              GlySerProGlyAla.ProGlyThrProGlI 298

1436 TGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGTGGGACAGGCATGTG 1387
     |||||| ::: |||||              |||||
298       yGluGly.....GlnGlnHisHisLeuGlyLeuGlyGlyAlaAlaLysGlnAlaGlyA 313
```

PI   Cappello J, Ferrari FA;
DR   WPI; 98-387004/33.
PT   Recombinant collagen-like polymers - useful for making gels, films,
PT   fibres, etc.
PS   Example 7; Column 50; 93pp; English.
CC   This sequence represents a unnatural collagen-like polymer
CC   of the invention. The products may be used as films, fibres, moulded
CC   objects and admixed with other natural or synthetic polymers or coatings
CC   on fibres, films, labware or other surfaces, e.g. prosthetic devices. The
CC   polymers may be used for binding a wide variety of specific binding
CC   materials, as catalytic substances (where the amino acid sequence may
CC   specifically chelate a wide variety of elements), as purification media,
CC   composites, laminates or adhesives. They may also be combined with
CC   inorganic or organic materials such as carbon fibres, nylon fibres,
CC   nitrocellulose, etc., as flask coatings or in synthetic matrices for the
CC   growth and study of cells, as affinity columns or as supports for
CC   biological materials. The polymers have collagen-like properties, but may
CC   be easily expressed in micro-organisms in high efficiency. The new
CC   sequences can be tailored to give the desired properties.
SQ   Sequence 762 AA;

alignment_scores:
        Quality: 393.00   Length: 852
        Ratio: 1.059   Gaps: 59
Percent Similarity: 43.545   Percent Identity: 30.516

alignment_block:
US-09-030-606-110/rev x W57666

Align seg 1/1 to: W57666   from: 1 to: 762

```
2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAACCCTTCTCTAGG 2298
     |||:::||||                ::||||:::||:|
  76 GlySerProGly.............AlaProGlyThrProGlyGluGl 87

2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATC 2251
     |    ||   ::||||||||||     :::||:||:|
  87 yGlnGlnHisHisLeuGlyLysGlnAlaGlyAlaGlySerP 104

2250 CACCTGCA........GAGTCCCGCCATTCAGTGCATGGAGCCCTTCT 2210
     ||  |||         ::|||:||||::|
 104 roGlyAlaProGlyThrProGlnGlyLeuProGlySerPro... 119

2209 GGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTC 2160
     |||   |||            |||||||||   ::|:||||
 120 GlyAlaPro.........GlyThrProGlyPro...GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGAGAGAGGACGGCCCCAGCCCCCCAG 2110
     ||||||||                 ||||||:||
 131 oGlySerPro.............GlyAlaProGlyThrProGlyProGlnG 144

2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT......GGCAGCAGAGAG 2066
     ::::||   ::|||  :::|||  ||   |||      |||:::::
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159

2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCCAGCCCCA 2016
     |||||||::::||||
 160 GlnHisHisLeuGly... 164

2015 TGGGGCTAACAGGAGGCGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCAC 1966
     ||||||||||||||:|||||:|  |||  |||||||||
 165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180

1965 CCCAATGTGCTGGAAGTTTTCTACCGCTGAGTATTTGCCAAGTCGCTCTT 1916
     ||||                              :::
 180 rPro.......................GlyP 183

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGCAGTAAATGGCGACCCAGGCCTG 1866
     ::|||||                     ::|||||
 183 roGlnGlyLeuPro................GlySerProGlyAla 192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAG 1816
```

```
581 ProGlyGluGlyGlnGlnHisHisLeuGlyLysGlyAlaLysGlnAlaGlyAs 597
         :::::       ||||||
581 ...GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATGAAGGGCCG 539
         |||||||  |||:::||        |||  :::||
597 pValGlySerProGlyAlaProGlyThrProGlnGlyLeuProGlyLeuProG 614
       ||||||  |||||||||
538 GCGGCGGCCATAGCGTCCACGCGCCAGTGGTCACTGGCTGCCTAGGAGCG 489
       ||      |||||
614 ly.........................SerProGlyAla 618
488 GGACACAGACAGGCCCAGCACTGACCAATGCCCAGCACCATGGTCATG 439
       |||||||
619 ProGlyThrProGlyPro......................... 624
438 AACTTCTCCCTCACCCCCACTTCCAGCAGCAGAGGCGGCACATAGGTGAT 389
         :::   ||||::::
625 ...GlnGlyLeuProGlySerProGlyAla..................... 633
388 GCCTGCGGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339
       :::::||||||  |||  |||:::
634 .......ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647
338 AGAGCTGGGCTTTCCGGT........................... 321
       |||   |||
648 AlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyLysGlyAlaLy 664
       :::   ||||||||||||||
320 ...GCCGCAGCAGGCGGCTCACCCACCACCAGCCTCTGGACCATAGTGGGCCA 275
       |||:::          ||||
664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680
       |||::::          |||||
274 GGCGGG...TAGGGCTCAGGGGGCGGTCAGGCACTCCAGAACTGCTTCG 228
       |||         ||||||||
681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlI 697
         ::|||||          |||||||
227 TCTCGGCTCTGCTCCAGAAGCTGCGGCCTTCTCCTCCTTGCTGCCGCCAAC 178
       ||||:::||          ||:::||||
697 yLeuProGlySerProGlyAlaProGlyThrProGly.............. 709
       ||||||||||||||||
177 TGCCTAGGAATCAGCCAGGCGCCCATTCTGCCCAGCCCCTTTGGTGCCGGT 128
         :::::||||:::          |||          |||
710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724
       ::|||||::||          |||   |||
127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTG 78
         ::::::|||||          ||:::||
725 GluGlyGlnGlnHisHisLeuGlyLy.........GlyAlaLysGlnAl 737
       :::::|||          |||||   |||
77 GGGACACGTCTCATCA.........CTCAGATCCTGGCCGA 46
       ::::::||||          ||||||||
737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
```

seq_name: A_Geneseq_36:W57666

seq_documentation_block:
ID   W57666   standard; peptide; 762 AA.
AC   W57666;
DT   27-AUG-1998 (first entry)
DE   Collagen-like polymer.
KW   Collagen-like polymer; synthetic polymer; fibre coating;
KW   prosthetic device; catalytic substance.
OS   Synthetic.
PN   US5773249-A.
PD   30-JUN-1998.
PF   02-MAY-1996; US5773249-A.   642255.
PR   02-MAY-1996; US-642255.
PR   04-NOV-1986; US-927258.
PR   29-OCT-1987; US-114618.
PR   09-NOV-1988; US-269429.
PR   06-NOV-1990; US-609716.
PR   12-NOV-1991; US-791960.
PR   05-NOV-1992; US-972032.
PR   22-DEC-1995; US-577046.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

```
193 ProGlyThrProGlyGln...............................                               199
     |||||
                                                                                    |||

1815 CCCATAAACAGGATGGGCGCCACCTGGGACACGCAGGAAGGCACTATCCAG                             1766
     |||||                   |||||::::||     |||         |||
200  .................GlyLeuProGlySerProGlyAla......ProG                            209

1765 GATGGCGAGGTCCAGGCAGATGCCCCGGC...........CCGGAACCA                              1728
     || |||||          ||||||            |||  :::
209  lyThrProGlyProGlnGlyLeuProGlySerProGlyAla.GlyTh                               225

1727 CCCTGGCCTCGGTGGGCTCACCCACCACCACACGTACGGAGACATCACAG                             1678
     |||||||  |||         ::||||||||  |||  ::::::||||
225  rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlaLysGln                             239

1677 GCA.........GAGGCCCGCAGAGCGCGGTGGAGGTGGGAGGCAGGCC                              1637
     |||        :::||    ||  ||  |||||||||
240  AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl                            256

1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTAG                             1587
     |||||  ::||                  ||||||||||||
256  yLeuProGlySerPro..............GlyAlaProGlyThr.                                266

1586 GGCCTGGCCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCA                            1537
     |||||  ::: |||  ||       |||
267  ..ProGlyProGlnGlyLeu.ProGlySerPro.............                                276

1536 GTGTCCCCTCGTATTTGGCGAACACCTGCTTCTCCCGGTGGTAGAG                                 1487
     |||||||||   ::::||||  |||||     ||||||
277  ...............GlyAlaProGlyThrProGlyProGlnG                                    286

1486 GGAGGCCAGTGTGTGTAGGGCAGGATCTCGAGGGCTGAGAAGGTGAACCCGG                           1437
     ||  |||      ||||| |||        |||  ||||||
286  lyLeuPro.........GlySerProGlyAla.ProGlyThrProGl                               298

1436 TGAGGGCGGGCTGAAGCTGTCACCACGGCCACACTGTGGGACAGGCATGTG                            1387
     ||||||||        :::  ||||||
298  yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA                             313

1386 GCACCGGCAG..............CCACAGGGAAAGCTG                                        1362
     ||||||||                   ||||||||  :: |||
313  spValGlySerProGlyAlaProGlyThrProGlyGly...Leu                                  328

1361 CCACACTGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCGCGG                             1312
     ||| |||                   :::||||||:::
329  ProGlySerPro..................GlyAlaProGlyTh                                   337

1311 TCCATGACCAGGAGAAGACCAGGGAGATGGCGCACTGCAGGAACA......                            1266
     |||| |||:::  |||||||  ::: ||||
337  rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG                            354

1265 .................................GCCCCAGGCTGCCCATCCGAA                         1245
     ||||||||| |||::||||
354  lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu                            370

1244 CGCCTTCATCATAGTGTCTCCGGCCTCCGGCCTCAGCTCTGGGC                                   1195
     ::::|||  :::::||     |||:::        :::
371  GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe                            387

1194 A......CGCCTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG                              1151
     |||:::||||||     ||| |||  |||::::||
387  rProGlyAlaProGlyThr...ProGlyProGlnGly.............                            398

1150 CGTGAAGGTCATGAGTGCCATCCATCAGCTGCAGCGTCAGCCACGAAGAGCC                           1101
     ::::||:::|||       ||      |||
399  .............LeuProGlySerProGlyAlaProGlyThrProGly                            410

1100 GGCCAGGGTGCGGGGCATGCGGGCAGCAGCTGGTCAGCCCGGGGAAGC                               1051
     |||                ::::: |||:::  |||
411  ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl                            427

1050 AGGGCGCCCAGGTTCC...GGAAAGCCAAGCGAGCGGGCCCGCAT.GGACAGC                          1005
     |||  ..|||||||||   |||:::      |||||||
427  nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG                            444
```

```
427  nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG                            444
     |||||

1004 AG.........TGGGCGCGACAAGGAGGAGGGGCCGACAGCCCTTCTGCTGGC                          964
     ||        |||  |||  ||||||||||||
444  lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly                           460

963  TCGGTGGGG................................CCCAGCGCTGC                           944
     ::: |||                                  |||::::||:::|
461  AlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.P                            477

943  CTCCTCAGCCACCAGCAGTGTGGCTCCTACCCAGGTGAGGAGAATGAGGG                            894
     ||  |||                   |||  ::::::||||
477  rGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly                             493

893  TGAGCAGGCCAAAGAGCCACTCCT..............CCTGGGTGCCC                             859
     |||                   :::||    ||
494  ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr                            510

858  AGGTAGGGGCCAGGGCACTGGTGTCCCAGTCAATGCCAGCAGGA.....                             813
     |||::::|||   |||     ||||
510  oGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspV                            527

812  ..GGTAGCCCAGGCAGCCCCCAAGACTGATCATGAGGCATAGACAGAGT                             765
     |||||||||||||::|||                       :::
527  alGlySerProGlyAlaPro.............................Gly                          534

764  AGGCCTGGCGACAGTGT...CCGGGT...CCCGAAGAGGTCAGAGAGC                              721
     |||||         ::::|||    ||    |||||
535  ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr                            551

720  AGGGCCCCAGTGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCC                              671
     |||||                  :::||:||:::|||:::
551  oGlyProGlnGlyLeuProGlySerProGly......AlaProGlyThrP                            566

670  CACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATCCGGCACA                             621
     || |||                   :::||||||||
566  roGlyProGlnGlyLeuProGlySerProGlyAla......ProGlyThr                            580

620  GCGAGCCCTGCTAGCCAGCCAGCCGGCCCCTTGGGATGGAGAAAGA.....                           582
     :::                     |||||
581  ProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAs                            597

581  ....GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATGAAGGGCCG                            539
     ||||||       |||:::|||
597  pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG                            614

538  GCGCGGGGCATAGCTCCACGCGTCCCAGCAGTGTCACTGGCTGAGCGGACG                           489
     ||:::|||||       ||| |||    |||||:::::||
614  ly...................................SerProGlyAla                            618

488  GGACACAGACCAGGCCCAGCACTGGACCAATGCCCAGCCACCATGGTCATG                           439
     ||||||||
619  ProGlyThrProGlyPro...............................                            624

438  AACTTCTCCTCTACCCCACCTTCCAGCAGCAGAGGCGGCACATAGGTGAT                            389
     :::  ||||||   |||:::||
625  ....GlnGlyLeuProGlySerProGlyAla......................                         633

388  GCCTGCGGCCAAACACACTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA                             339
     |||:::||||||    ||| |||  |||:::::
634  .......ProGlyThrProGlyProGlnGlyLeuProGlySerProGly                            647

338  AGAGCTGGGCTTTCCGGT.............................                               321
     |||:::|||
648  AlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLy                            664

320  .....GCCCAGCAGGCGGCTCACCCACCAGCCCTCTGGACCCATAGTGGGCCA                         275
     |||||       |||::||         ||||||
664  sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro                           680

274  GGGCGG....TAGGGCCAGGGGCCGTTCAGGCACTCCAGCAACTGCTTCG                            228
     |||       |||  |||||||||
681  GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl                           697
```

```
 227 TCTCGGCTCTGCTCCAGAAGCTGGGGCTCCTTCCTCCTTGCTGCGCCAAC 178          87 yGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 104
         |||   :::::||| |||  ::||||| |||                             2250 CACCTGCA........GAGTCCCGCCATTCCAGTGCATGGAGCCCCTTCT 2210
 697 yLeuProGlyGlySerProGlyAlaProGlyThrProGly............. 709          || ||       ::::||| |||:||||| |||
                                                                   104 roGlyAlaProGlyThrProGlnGlyLeuProGlySerPro...         119
 177 TGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTGGTGCCGGT 128
         |||::::::||| :||||||:: :||  |||  :||||||||         2209 GGCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCTCCAGTC 2160
 710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724          |||  |||        |||     ::|:|||||
                                                                   120 GlyAlaPro.......GlyThrProGlyPro...GlnGlyLeu....Pr 131
 127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTG 78
         :::::||||     |||::||                            2159 AGGCAGCCCTAGAGACTGGGAGAGGAGAGGGACGCCCCAGCCCCCAG 2110
 725 GluGlyGlnGlnHisHisLeuGly...........GlyAlaLysGlnAl 737          |||||||||            |||||| |||  :
                                                                   131 oGlySerPro...........GlyAlaProGlyThrProGlyProGlnG 144
  77 GGGACACGTCTCATCA........CTCAGATCCTGGCCGA 46
         |||  ||::::|||  ::: ||||||||||||                2109 CTGTGCAGCTACGCCACCTCAGCACCAGGGT........GGCAGCAGAGAG 2066
 737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750          :::::   ||||   |||:::||  ::::|| |||:|
                                                                   144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluProGlyGln 159

seq_name: A_Geneseq_36:W49713                                     2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCCAGCCCCA 2016
                                                                         |||::::::||||
seq_documentation_block:                                          160 GlnHisHisLeuGly.................................. 164
ID  W49713 standard; Protein; 762 AA.
AC  W49713;                                                      2015 TGGGGCTAACAGGAGCGGGGAGCTGGACCCAGTGAGGCAGGCCCTCCAC 1966
DT  12-OCT-1998  (first entry)                                         |||||||::::::|||:||:||:||||:|||  ||| |||::||
DE  Protein polymer adhesive substrate PPAS1-A.                  165 .GlyAlaLysGlnAlaGlyAspValGlySerPro.GlyAlaProGlyTh 180
KW  Protein polymer adhesive substrate: PPAS1-A; sealant;
KW  wound healing; CLP 3.7; collagen: human.                     1965 CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGCCAAGTCGCTCTT 1916
OS  Synthetic.                                                         |||                              ::|
OS  Homo sapiens.                                                180 rPro...............................GlyP 183
PN  US5773577-A.
PD  30-JUN-1998.                                                 1915 GTCAAATACTACCTGTGTGAGCAAAGTAAATGGCAGCACCCAGGCCTG 1866
PF  03-MAR-1994; 205518.                                              |||  ||| |||||
PR  02-MAR-1995; US-397633.                                      183 roGlnGlyLeuPro...............GlySerProGlyAla 192
PR  03-MAR-1994; US-205518.
PA  (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.                    1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGAG 1816
PI  Cappello J;                                                        |||||||      |||
DR  WPI: 98-387091/33.                                           193 ProGlyThrProGlyProGln.......................... 199
PT  New recombinant protein polymers - containing naturally occurring
PT  repetitive units for crosslinking by enzymes, useful as medical  1815 CCCATAAACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCACTATCCAG 1766
PT  adhesives and sealants, depots and matrices                      |||||||||::: |||
PS  Example 4: Column 26; 70pp; English.                         200 ...............GlyLeuProGlySerProGlyAla........ProG 209
CC  Protein polymer adhesive substrate PPAS1-A comprises 10 repeats
CC  of a monomer (see W49712) consisting of human fibrin gamma chain 1765 GATGGGCGGAGGTCCAGGCAGATGCCCCGGC..........CCGGAACCA 1728
CC  peptide (aa421-437) inserted into a structural backbone of collagen  ||  |||||||        |||  |||  :::
CC  repeat motifs (see W49711). The 68 kDa PPAS1-A polymer has been   209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225
CC  expressed in Escherichia coli HB101 transformants using plasmid
CC  pPT0321. Claimed recombinant protein polymers are capable of  1727 CCCTGGCCTCGGTGGGCTCACCCACCACCACGTACGGAGACATCACAG 1678
CC  covalent crosslinking by enzymatic reaction to form products which  |||||| |||           ::|||||| |||         |||
CC  set quickly and have good adhesive properties and high strength.  225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlaLysGln 239
CC  CLP (collagen like protein) polymers are extremely soluble in water,
CC  allowing protein solutions of over 10 wt.% to be formed while  1677 GCA........GAGGCCCCAGAGCGCGGGTGGAGGTGGAGGACGAGGCC 1637
CC  maintaining good flow properties. CLP polymers have good adhesion   |||          :::|||      |||  ::
CC  to hydrophilic surfaces and therefore may adhere well to tissue.  240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 256
CC  The fibrin peptide inserts of PPAS1-A provide sites for Factor
CC  XIIIa cross-linking.                                          1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGAAGGAGGAGCTCCAGGCTTAG 1587
SQ  Sequence  762 AA;                                                  |||||||| :::|||                |||||||||||||
                                                                   256 yLeuProGlySerPro.................GlyAlaProGlyThr. 266
alignment_scores:
        Quality:  393.00                    Length:  852           1586 GGGCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCA 1537
        Ratio:  1.059                        Gaps:  59                   |||||| ::: |||     |||  ||| |||
Percent Similarity:  43.545        Percent Identity:  30.516        267 ..ProGlyProGlnGlyLeu.ProGlySerPro............... 276

alignment_block:                                                   1536 GTGTCCCCTCGGTGATTTGGGCAGGAACACCTGCTTCTCCCGGTGGTAGAG 1487
US-09-030-606-110/rev x W49713  ..                                    |||:::|||:|||:||:||||  ||| ||| ::::|
                                                                   277 ......................GlyAlaProGlyThrProGlyProGlnG 286
Align seq 1/1  to: W49713  from: 1  to: 762
                                                                   1486 GGAGGCCAGTGTGTAGGCAGGCCAGGATCTGCAGGGCTGAGAAGGTGAACCCGG 1437
2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGG 2298          ||  |||            ::|||:::||  ||
         |||:::||||||     ||||||| ::||||||:::: ||          286 lyLeuPro..........GlySerProGlyAla.ProGlyThrProGl 298
  76 GlySerProGly.................AlaProGlyThrProGlyGluGl 87
                                                                   1436 TGAGGGGCGGCTGAAGCTGTCACCACGGCCACACTGTGGACAGGCATGTG 1387
2297 TGTGTCT...CAACTAGGAGGCTAGCTGTGTTAACCCTGGTAATC 2251          ||||||              ::: ||||||
         |  :::  |||||||||            ::: |||:||:|::|       298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 313
```

```
 620 GCAGCCCTGCTAGCCAGCCGGCGCCCTTGGGATGAGAAAGA.......... 582
         :::   ||||||                    :::::   ||||||
 581 ProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAlaAs  597

 581 ...GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATGAAGGGCCG  539
       ||||||  |||::||| |||         ||:||
 597 pValGlySerProGlyAlaProGlyThrProGlnGlyLeuProG  614

 538 GCGGCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGAGCG  489
     ||                                :::   ||||||
 614 ly..............................SerProGlyAla  618

 488 GGACACAGACCAGGCCAGCACTGACCAATGCCCAGCACCATGGTCATG  439
                ||||||||
 619 ProGlyThrProGlyPro................................  624

 438 AACTTCTCCTCTACCCCCACTTCCAGCAGCAGAGGCGCACATAGGTGAT  389
                                        :::   |||::|||
 625 ....GlnGlyLeuProGlySerProGlyAla..................  633

 388 GCCTGCGGCCAAACACACTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA  339
                                        |||::||||| |||::::
 634 .......ProGlyThrProGlnGlyLeuProGlnGlySerProGly  647

 338 AGAGCTGGGCTTTCCGGT..............................  321
       |||            |||
 648 AlaProGlyThrProGlyLyGluGlyGlnGlnHisHisLeuGlyGlyAlaLy  664

 320 ....GCCGCAGCAGGCGGCTCACCCAGCCTCTGGACCATAGTGGGCCA  275
         |||::|           ||||||||        |||    |||||
 664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro  680

 274 GGGCGG...TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAACTGCTTCG  228
       |||                                |||||
 681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlnGlI  697

 227 TCTCGGCTCTGCTCCAGAAGCTGCGGCCCTCTCCCTTGCTGCCGCCAAC  178
             |||     |||        |||  |::
 697 yLeuProGlySerProGlyAlaProGlyThrProGly..............  709

 177 TGCCTAGGATCAGCACCAGGCGCCCATTCTGCCAGCCCCTTGGTGCCGGT  128
     |||:::::  ||||||:::          :::     |||
 710 .ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly  724

 127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGTGCTGGGGCCACCTCAGTG  78
     |||||||||::||| |||                 |||||
 725 GluGlyGlnGlnHisHisLeuGly..............GlyAlaLysGlnAl  737

  77 GGGACACGTCTCATCA..............CTCAGATCCTGGCCGA  46
     ||| |||:::||                 :::  ||||||||||
 737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg  750
```

```
seq_name: A_Geneseq_36:R80330

seq_documentation_block:
ID   R80330 standard; Protein; 682 AA.
AC   R80330;
DT   18-APR-1996  (first entry)
DE   Protein polymeric adhesion substrate 1-C.
KW   Pendent group; repeating unit; enzyme recognition site; sealant; fibrin;
KW   enzymatic cross-linking; biocompatible material; structural integrity;
KW   medical adhesive; wound closure; tissue repair; transglutaminase.
OS   Synthetic.
PN   WO9523611-A1.
PD   08-SEP-1995.
PF   03-MAR-1995; U02728.
PR   03-MAR-1994; US-205518.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
DR   WPI: 95-320413/41.
PT   Protein polymers comprising repeating units and sequences  - capable
PT   of enzyme-catalysed covalent bond formation useful as a
PT   biocompatible material for wound closure and tissue repair
```

```
1386 GCACCGGCAG.............................CCACAGGGAAAGCTG 1362
         :::|||                            ||||||||    |||
 313 spValGlySerProGlyAlaProGlyThrProGlnGly...Leu 328

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCCACCAGCCGG 1312
     |||         :::||||||::::
 329 ProGlySerPro...................GlyAlaProGlyTh 337

1311 TCCATGACCAGAGAGAAGAACCAGGGAGATGGCCACTGCCAGGAACA....1266
     ||||||   ||||||           |||::::   ||||||
 337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354

1265 ......................GCCCAGGCTGCCCATCCGAA 1245
                          ||||||||  |||::::|||
 354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTGTCTCCGGGCCTCCGGTGCCCGGCTCAGCTCTGGGC 1195
         :::||   :::::||| ||       ::
 371 GlyGlnGlnHisHisLeuGlyLyAlaLysGlnAlaGlyAspValGlySe 387

1194 A......CGCCCTGGTACAGCCCCTGCCCACGAAATCCGTGTAAAACAG 1151
     |      ||||||||  ||| |||:::::
 387 rProGlyAlaProGlyThr..ProGlyProGlnGly............ 398

1150 CGTGAAGGTCATGATGCCATCCAGCTGCCACAGCTCAGCCACGAAGAGCC 1101
     |||::::|||    |||
 399 ............LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGGCAGGTGCGCGGGCATGCGCGAGCACAGCTGGTGCAGCCGGGGAAGC 1051
     |||      ::     |||::|      |||
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGCGCCCAGGTTCC...GGAAAGCCAGGCGGGCCCGGCAT.GGACAGC 1005
     |||         |||   ||| :::  |||::::
 427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444

1004 AG........TGGGCGACAAGGAGGGGGCCGACAGCCCTTCTGCTGGC 964
     ||         |||    ||| :::      |||
 444 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 460

 963 TCGGTGGGG....................CCCAGCCGCTGC 944
     |||                          |||:::|||
 460 ...................................Gly 534

 943 CTCCTCAGCCACCACAGTGTGGCTGCTACGCAGTGAGGAAGATGAGGG 894
     |||         |||          |||
 461 AlaProGlyThrProGlyGlnGlyLeuProGlySerProGlyAla.P 477

 893 TGAGCAGGCCAAAGAGAGGCACTCCT..........CCTGGGTGCCC 859
     ||    :::|||  ||                    |||   |
 477 roGlyThrProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 494

 858 AGGTAGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA..... 813
     |||::::|||                        :::   ||||||
 494 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510

 812 ..GGTAGCCCAGGCCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGT 765
     |||||::|||||||||||||::||                  :::
 510 oGlyGlyGlnGlnHisHisLeuGlyLyAlaLysGlnAlaGlyGlnAlaGlyAspV 527

 764 AGGGGCCTCCAGTGGAGTGAGCAGCACACCTGGCCACAGAAGTCCAGCAGAGC 721
     ||||||   ||||||::||||| |||
 527 alGlySerProGlyAlaPro...CCGGGT...CCCGGAAGAGGTCAGGAGAGC 721

 720 AGGGGCTCCAGTGGAGTGAGCAGTCCACAGGGCCTGGGATCCGGGCCACA 621
     ||||||   :::::||||| |||||||
 535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551

 670 CACGGCCAGGATGAGCAGTGCCAGTCCAGGGGCTTCGGGATCCGGGCCACA 621
     ||  |||                 |||||||||
 551 oGlyProGlnGlyLeuProGlySerProGlyAla.....AlaProGlyThrP 566

 566 roGlyProGlnGlyLeuProGlySerProGlyAla.....ProGlyThr 580
```

```
PS  Example 4; Page 47; 138pp; English.
CC  The amino acid sequence of the protein polymeric adhesion substrate
CC  (PPAS) 1-C. The protein comprises 10 repeats of the PPAS1-C monomeric
CC  repeat (R80329) which consists of the CLP 3.7 gene encoded sequence
CC  (R80321) linked to the human fibrin cross-linking substrate peptide
CC  93.4 sequence (R80317). The polymers can be used in biological systems
CC  where in situ formation of a biocompatible material with structural
CC  integrity is required e.g. as medical adhesives and sealants or for wound
CC  closure or tissue repair.
SQ  Sequence 682 AA;

alignment_scores:
        Quality:  370.50         Length:    831
        Ratio:    1.015          Gaps:       59
Percent Similarity:  43.923      Percent Identity:  29.964

alignment_block:
US-09-030-606-110/rev x R80330

Align seg 1/1  to: R80330  from: 1  to: 682
```

```
2314 AAAAACCCTTCTCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     ::|||||||   ||||||::|||||||::
  12 GluAsnPro.....GlyValThrGlnLeuAsn...........ArgLe   22

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATTCCAGTGCATGGAGCC 2215
     |:||   |||  |||||:::||| :::||| :::||: :::|||
  22 uAlaAlaHisProProPheAlaSerAspProMetGlyAlaProGlyThrP 39

2214 CTTCT............GGCCTCCCTGTATAAGTCCAGACTGAAACC     2180
     ||:|:            |||  ||| |||       ::|||
  39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55

2179 CCCTTGGAAGGCCTCCAGTCAGCAGCCCTAGAGACTGGGGAGAGGAG    2130
     ||| ::||||||   ||||||||                 :::||
  56 Pro...GlnGlyLeu..ProGlySerPro...........GlyAlaPr  66

2129 AGGGACGCCCCAGCTGTGCAGCTACGCACCTCAGCA......G       2086
     |||||||||  |||  ::||||           ::|||
  66 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 83

2085 CACAGGGTGGCAGCAGAGACCCACATTACTTTGGCAGCAA..........  2046
     ::  |||||||::::::::  ::|||    ||||||
  83 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99

2045 ..........CAGAAACTGGGCGGCCAGCCCGCAGCCCATGGGG       2011
     |||  ||| :::||||||::::||||
 100 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 116

2010 CTAACAGGACGCGGGAGCTGGACCCAGTCAGGCAGGCCCTCCACCCCAA   1961
     ::||||| |||:||::::|||
 116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro. 132

1960 TGTGCTGGAAGTTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTTGTCAA  1911
     :::  |||              ::: |||
 133 ..............GlyProGln 135

1910 ATACTACCTGTGTAGCAAAGTAAAATGGCGACCAGGACCGCTGCGGCA   1861
     ||||||              ::||||||
 136 GlyLeuPro.............GlySerProGlyAlaProGl 145

1860 GACACCATATAGGCAGTGACGACTGGCTGAGCTGACAATGGAGCCCAT    1811
     |||||  ::||||||  :::|||   ||||||
 145 yThrProGlyGlyAlaLysGlnAlaGlyAspValGly.....SerPro.  159

1810 AAACAGGATGGGGCACCTGGGACACGCAGGAAGGCAC...........   1773
     ||| ||||||||||  ||| ||||||       :::|||
 160 .........GlyAlaProGlyThrProGlyProGlnGlyLeuProGly  172

1772 ........TATCCAGGATGCCAGGTCCAGGCAGATGCCCCGGC.....   1737
     |||||||            |||||
 173 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 189
```

```
1736 ......CCGGAACCACCTGGCTCGGTG......GGCTCACCACCA     1701
     |||   |||||||||||   ||||||||| :
 189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 206

1700 CCACACGTACG...........GAGAACATCACAGGCAGAGGCCCCG   1666
     :|::          |||||||::||| :::|||
 206 laProGlyThrProGlyGlyAlaLysGlnAlaGlyAspValGlySerPro 222

1665 CAGAGCGCCGGGTGGGAGGTGGGAGCAGGCAGCCCTCCAGCACCCACGTG  1616
     :::  ||| :::  |||      |||||||  :::|||
 223 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 239

1615 TCCATTAGGGAAGGAGCTCCAGGTTAGGGCTGGCCAGGAAGCTGGTCA   1566
     :|||  ||||||||       ||||||   ::: :::
 239 aProGlyThrProGlyProGlnGlyLeu..ProGly.SerProGlyAla 254

1565 TCAGGCTGTCGTCCTCACTGCTAGCACCTCCAGTGCCCCTCGGTATTTGGGC 1516
     |||   |||||||::::::|||          ::||: |||
 255 ProGlyThrProGly..............................ProGlnGlyLeu 263

1515 AGGGAACACCTGCTTCTCCCGGTAGGAGGCCA.....GTGTGTA        1472
     ||::|||||||||||||||||| ||| |||
 264 .....ProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysG 278

1471 GGGCAGGGATCTGCAGGGTGAGAGGTGAACCCGGTGAGGGCGGCTGAAG   1422
     ::|||||||:::::||
 278 lnAlaGlyAspValGly............................ 283

1421 CTGTCACCACGGCCACACTGTGGGACAGGCATGTGGCACCGGCAGCCACA  1372
     ||||||                       ::||||
 284 ...SerProGlyAlaProGlyThrProGly..................ProGl 294

1371 GGGAAAGCTGCCACACTGGCCAAATAGACTGCTGCAGTGCCGAATCGCTG  1322
     ||||||              :::
 294 nGly...LeuProGlySerPro..................GlyA 302

1321 CACCAGGCCGGTCCATGACCAGGAGAGAGACCAGGAGATGGCCACTGCA   1272
     ||||::: ||| ||| ::::
 302 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 318

1271 GGAACA...........................GCCCCAGGCTG      1255
     ||||||                           |||||||
 319 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335

1254 CCCA...TCCGAACGCCTTCATCATAGTGTCTTCCGGGCCTCCGGTGCCCGG 1208
     |||    :::|||||||::: :::||:::
 335 rProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG 352

1207 CTCAGCTCTGGGCACGC.......CCTGGTACAGCCCCTCGCCCACGAAAT 1164
     ::: |||||||::::  |||    :::
 352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368

1163 CCGTGTAAAAACAGCGTGAAGGTCATGAGTGCCATCAGC.........   1125
     |||                ::|||:::|||
 369 ProGlyProGln.........LeuProGlySerProGlyAlaPr 381

1124 .TGCACAGCTCAGCCACCAAGAGCCCGCAGGGTGCGGGCCATGCGGCA   1076
     |||   :::  |||   :::|||
 381 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 398

1075 GCACAGTGGTGCA...........GCCGGGAAGCAGGCGGCCCAGGTTCC 1035
     |||::||||||           |||     :::||
 398 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 414

1034 GGAAAGCCAAGCGGGCCCCGGCCATGGACAGCAGTGGGGCGACAAGGAGGG 985
     ||| :::|||            :::||: :::|||
 415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431

984 .GCCGACAGCCCCTTCTGCTGGCTCGGTGGGCCCCA..GCGCTGCCTCCTC 938
     |||||||             |||||
 431 rPro.................................GlyProGlnGlyLeuProGlyLys 440
```

```
                                      |||::::|||                    :::  |||||||||||
                            659 AspValGlySerProGlyAlaMet.AspProGlyArg 670

seq_name: A_Geneseq_36:W57670

seq_documentation_block:
ID   W57670  standard; peptide; 682 AA.
AC   W57670;
DT   27-AUG-1998  (first entry)
DE   Collagen-like polymer.
KW   Collagen-like polymer; synthetic polymer; fibre coating;
KW   prosthetic device; catalytic substance.
OS   Synthetic.
PN   US5773249-A.
PD   30-JUN-1998.
PF   02-MAY-1996;  642255.
PR   02-MAY-1996; US-642255.
PR   04-NOV-1986; US-927258.
PR   29-OCT-1987; US-114618.
PR   09-NOV-1988; US-269429.
PR   06-NOV-1990; US-609716.
PR   12-NOV-1991; US-791960.
PR   05-NOV-1992; US-972032.
PR   22-DEC-1995; US-577046.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J, Ferrari FA;
DR   WPI; 98-387004/33.
PT   Recombinant collagen-like polymers - useful for making gels, films,
PT   fibres, etc.
PS   Example 7; Column 53; 93pp; English.
CC   This sequence represents a unnatural collagen-like polymer
CC   of the invention. The products may be used as films, fibres, moulded
CC   objects and admixed with other natural or synthetic polymers or coatings
CC   on fibres, films, labware or other surfaces; e.g. prosthetic devices. The
CC   polymers may be used for binding a wide variety of specific binding
CC   materials, as catalytic substances (where the amino acid sequence may
CC   specifically chelate a wide variety of elements), as purification media,
CC   composites, laminates or adhesives. They may also be combined with
CC   inorganic or organic materials such as carbon fibres, nylon fibres,
CC   nitrocellulose, etc. as flask coatings or in synthetic matrices for the
CC   growth and study of cells, as affinity columns or as supports for
CC   biological materials. The polymers have collagen-like properties, but may
CC   be easily expressed in micro-organisms in high efficiency. The new
CC   sequences can be tailored to give the desired properties.
SQ   Sequence  682 AA;

alignment_scores:
     Quality:      370.50          Length:     831
     Ratio:        1.015           Gaps:        59
Percent Similarity: 43.923         Percent Identity: 29.964

alignment_block:
US-09-030-606-110/rev x W57670

Align seg 1/1  to: W57670  from: 1  to: 682

2314 AAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     ::::||||||    |||   ||||:::||:||::|:||||||||
  12 GluAsnPro.....GlyValThrGlnLeuAsn...........ArgLe  22

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATTCCAGTGCATGGAGCC 2215
     |::::              ||    :::: |::::: ::::|:|||
  22 uAlaAlaHisProProPheAlaSerAspProMetGlyAlaProGlyThrP  39

2214 CTTCT...............GGCTCCCCTGTATAAGTCCAGACTGAAACC 2180
     |:||:|            |||   |||     |||||||||||
  39 rGlyProGlnLeuProGlySerProGlyAlaProGlyThrProGly  55

2179 CCCTTGGAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGAGAGGAGGAG 2130
     |||  ::|||||||    |||||||||
  56 Pro...GlnGlyLeu..ProGlySerPro...........GlyAlaPr  66
```

```
937 AGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCA 888
    ::|||                          |||      |||
440 erPro...........................GlyAlaProGlyThrPro 447

887 GGCCAAAGAGGGCACTCCT.....CCTGGGTGCCCAGGTAGGGGCCCAGG 844
    |||  ::||||  |||   |||  ||||||:::||||||
448 GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProGl 463

843 GCACTGGTGTCCCAGTCAATGCCAGGCAGGA......GGTAGCCCAGGCA 800
    :::           |||||||||   ::|| |||||||||
463 y............GlyAlaLysGlnAlaGlyAspValGlySerProGlyA 476

799 GCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCGACAGT 750
    ::|||                        :::  |||||  :::
476 laPro........................GlyThrProGlyProGln 483

749 GGT...CCGGT...CCCGGAAGAGGTCAGAGAGCAGGGCCTCCAGTGGA 706
    |||  ||||| |||||     |||||||||||
484 GlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 500

705 GTGAAGCACACCTGCCACAGAAGTCCAGCAGCCCCACGCCCCAGGATGAG 656
    :::::::||||       :::::||||  |||
500 uProGlySerProGly......AlaProGlyThrProGlyProGlnGlyL 515

655 CAGTGCCAGCTCCCAGGGGCCTGGGATCCCGGGCACAGCAGCCCTGCTAGCC 606
    ::::||||||||||    :::  |||
515 euProGlySerProGlyAla......ProGlyThrProGly......Gly 527

605 AGCCGCCCCTTGGGATGAGAGAAAGAGCTCAGCAGGATGCCCAGGACAGT 556
    ::::||||    ||  |||||  ::::|||  |||::::||| |||
528 AlaLysGlnAlaGly..AspValGlySerProGlyAlaProGlyThrPr 543

555 G...CCCAGATGAAGGGCCGGCGGCGGCCCATAGCCGTCCACGCCAGTGGTC 509
    ||||:::                  |||
543 oGlyProGlnGlyLeuProGly.......................... 550

508 ACTGGCTGAGCCTAGGAGCGGGACAGACAGCCCAGCCACTGACCAA 459
    |||  |||||       ||||||     |||||||||
551 .......SerProGlyAlaProGlyThrProGlyPro.......... 560

458 TGCCCAGCACCATGGTCATGAACTTCTCCTCACCCCCACTTCCAGCAGC 409
    ||||  |||||   |||:::::|
561 ..................GlnGlyLeuProGlySerProGlyAl 569

408 AGAGGCCGGCACATAGGTGATGCCTGCGCGCCAAACACACCTCCAGGCCAAA 359
    |||::||||||| |||
569 a......................ProGlyThrProGlyProGlnG 577

358 GGTTAGGCAGGTTGACCAGCAAGAGCTGGGCTTTCCGGTGCCGCA...... 315
    |||  |||||  :::::     ||  |||:::
577 lyLeuProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysGln 593

314 GCAGGC........GGCTCACCCACAGCCTCTGACCATAGTGGCCAGGCG 271
    ||||||  |||||||||       ||| |||||
594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610

270 GG...TAGGGCTCAGGGGGGCCGTTCAGGCACTCCAGAAACTGCTTCGTCTC 224
    |||       |||   |||        |||||||||||
610 lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu 626

223 GGCTCGCTGCCAGAAGTGCGGGCCTCTCCCTTGCTGCCCGCCAACTGCC 174
    ||||||||  :::||    ||||||
627 ProGlySerProGlyAlaProGlyThr..GlyProGlnGly........Pr 638

173 TAGGAATCAGCCAGGCGCGCCCATTTCTGCCAGCCCTTGGTGCGGTCCAG 124
    |::::::||||||:::              ||  :::
638 oGlnGlyLeuProGlySer..........ProGlyAlaProGlyThrP 651

123 CTTCTCAGCCCCATGCTCAACACCTGCTGCTGTGGGCACCTCAGTGGGA 74
    :::          ||||| ||| |||||
651 rGly.....................GlyAlaLysGlnAlaGly 658

 73 CACGTCTCATCA.........CTCAGATCCTGGCCGA 46
```

us-09-030-606-110.rag

```
                      ||||||:::  |||  ||||||               ::
                      laProGlyThrProGlnGlyLeuProGlySerProGlyAlaPro  318
302

1271 GGAACA................................GCCCCAGCTG 1255
     ||||||
319  GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335

1254 CCCA...TCCGAACGCCTTCATCATAGTGTCCGGGCCTCCGGTGCCCGG 1208
     |||   :::         ::::|||:::||||| ::::|||  |||||
335  rProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG 352

1207 CTCAGCTCTGGCACGC.......CCTGGTACAGCCCCTCGCCCACGAAAT 1164
     ||||||||  :::           ||||| :::||||||
352  lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr  368

1163 CCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGC.......... 1125
     |||                            ::|||:::|||
369  ProGlyProGlnGly..........LeuProGlySerProGlyAlaPr   381

1124 .TGCACAGCTCAGCCACGAAGAGCCGCGCAGGTGCGGGCATGCGCGCCA  1076
     |||:::       :::::   |||::   |||||  |||
381  oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT  398

1075 GCACAGCTGGTGCA.....GCCGGGGAAGCAGGCAGGCGCCCAGGTTCC.. 1035
     |||  ::||||||    |||||||||| ::::||||
398  hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro  414

1034 GGAAAGCCAAGCGGGCCCGCATGGACAGCAGTGGGCGACAAGGAGGGG   985
     |||  ::|||||||||    |||  :::||
415  GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh  431

984  GCCGCAGCCCCTTCGTGGCCTCGGTGGGGCCCA...GCGCTGCCTCCTC  938
     |||        ::::|||||||                :::|||||
431  rPro..........................GlyProGlnGlyLeuProGlyLys  440

937  AGCCACCAGCAGTGTGGCTGCTACGCAGTGAGGAAGATGAGGGTGAGCA  888
     |||                                    :::|||
440  erPro.....................................GlyAlaProGlyThrPro  447

887  GGCCAAAGAGGCACTCCT.......CCTTGGGTGCCCAGGTAGGGGGCCAGG  844
     |||  ::|||  |||     ||||||    ::|||
448  GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProGl  463

843  GCACTGTGTCCCAGTCAATGCAGGCAGGA......GGTAGCCCAGGCA   800
     :::           ||||||||||||||
463  y.............GlyAlaLysGlnAlaGlyAspValGlySerProGlyA  476

799  GCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCACAGT  750
     :::|||           :::   ||||||
476  laPro.........................GlyThrProGlyProGln    483

749  GGT...CCGGGT...CCCGGAAGAGGTCAGAGAGCAGGGGCTTCCAGTGGA  706
     |||   |||||     ||||||
484  GlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe  500

705  GTGAAGCACACCTGGCCACAGAAGTCAGCAGCCCCACGCCCAGGATGAG  656
     ::::|||||||      :::|||||||    :::|||
500  uProGlySerProGly......AlaProGlyThrProGlyProGlnGlyL  515

655  CAGTGCCAGTCCCAGGGGCCTGGGATCCGGCACAGCAGCCCTGCTAGCC  606
     :::|||||||       ||||||||
515  euProGlySerProGlyAla.......ProGlyThrProGly......Gly  527

605  AGCCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCAAGGACAGT  556
     :::|||           ::  |||||||    |||||
528  AlaLysGlnAlaGly...AspValGlySerProGlyAlaProGlyThrPr  543

555  G....CCCAGATGAAGGCGCGGCGGCCATAGCGGTCCACCGCCAGTGTC  509
     |||:::       ||
555  oGlyProGlnGlyLeuProGly..........................   550

508  ACTGGCTAGCCTAGGACGGGGACAACAGACCCAGGCCACTGGACCAA   459
     |||       ||||||
508                                                     
```

```
2129 AGGGACGCCCAGCCCCAGCTGTCGAGCTACGCACCTCAGCA......G  2086
     ||||||||||    ||||    :::|||
66   oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT  83

2085 CACAGGGTGGCAGCAGAGAGCCACATTACTTGGCAGCAA..........  2046
     ::  ||||||::::::::      ||||||
83   hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro  99

2045 ..........CAGAAACTGGCGGCCAGCCCGGCAGCCCATGGGG       2011
     |||    :::||||||    :::|||:::|||
100  GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh  116

2010 CTAACAGGAGGCGGGGAGACTGGGACCCAGTGAGCGAGGCCCTCCACCCAA  1961
     :::||      :::  |||::||||||
116  rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro.  132

1960 TGTGCTGGAAGTTTTCTACGCTGAGTATTGGCCAGTCGCTCTTGTCAA   1911
     |||          :::  |||
133  .............................GlyProGln            135

1910 ATACTACCTGTGTAGTAGCAAAGTAAATGGCCACCAGACCCTGCCGGCA  1861
     ||||||
136  GlyLeuPro........................GlySerProGlyAlaProGl  145

1860 GACACCATATAGCAGTGACGAGACTGGCTAGCAATGGAGCCCAT       1811
     ||||||  :::||            ||||||
145  yThrProGlyGlyAlaLysGlnAlaGlyAspValGly......SerPro.  159

1810 AAACAGGGATGGGGCCACCTGGGACAGCAGGAGGCAC......       1773
     |||          :::
160  .............GlyAlaProGlyThrProGlyProGlnGlyLeuProGly  172

1772 ........TATCCAGGATGGCGAGGTCCAGGCAGCATGGCCCCGGC.....  1737
     :::          |||||||
173  SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr  189

1736 ............CCGGAACCACCCTGGCCTCGGGTG......GGCTCACCCACCA  1701
     |||  :::            |||||||
189  oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA  206

1700 CCACACTACG..........GAGACATCAGGCAGGAGGCCCG           1666
     ::   |||          :::||||||  ::|||
206  laProGlyThrProGlyGlyAlaLysGlnAlaGlyAspValGlySerPro  222

1665 CAGAGCGCCGGGTGGAGGTGGGAGGCCACTGCCTCCAGCCACCACGTG   1616
     :::  ||| |||
223  GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl  239

1615 TCCATTAGGGAAGGAGGAGCTCCAGGCCTTAGGGCCTGCAGGAAGCTGGTCA  1566
     :::|||    |||||||   ||||||
239  aProGlyThrProGlyProGlnGlyLeu...ProGly.SerProGlyAla  254

1565 TCAGGCTGTCCTCACTGCTAGCACCTTCCAGTGTCCCCTCCGTATTTGGGC  1516
     |||   |||          ||||||
255  ProGlyThrProGly............ProGlnGlyLeu..........  263

1515 AGGAACACCTGCTTCTCCCGGTGGTAGAGGGAGGCCA......GTGTGTA  1472
     |||::::::|||||||:::|||          |||||||
264  ........ProGlySerProGlyAlaProGlyThrProGlyAlaLysG   278

1471 GGGCAGGATCTGCAGGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAG  1422
     :::||||||:::::|||
278  lnAlaGlyAspValGly.................................  283

1421 CTGTCACCACGGCCACACTGTGGACAGGCATGTGCACCGGCCAGCCACA  1372
     ||||||     :::||||||    |||||
284  ...SerProGlyAlaProGlyThrProGly...................ProGl  294

1371 GGGAAAGCTGCCACACTGCCCAAATAGACTGCTCGAGTGCCGAATCGCTG  1322
     |||  |||||||        :::|||
294  nGly...LeuProGlySerPro.......................GlyA   302

1321 CACCAGCCCGGTCCATGACCAGACAGAGAAGACCAGGGAGATGCGCACTGCA  1272
     |||        ||||||
```

```
CC      tissue. The fibrin peptide 93.4 inserts of PPAS1-C provide lysine
CC      donor sites for Factor XIIIa mediated cross-linking.
SQ      Sequence    682 AA;

alignment_scores:
        Quality:   370.50          Length:       831
        Ratio:       1.015         Gaps:          59
Percent Similarity:  43.923   Percent Identity:  29.964

alignment_block:
US-09-030-606-110/rev x W49717

Align seg 1/1  to: W49717  from: 1  to: 682

2314  AAAAACCCTTCTCTAGGTGTGTCAACTAGGAGGCTAGCTGTTAACCCT  2265
                                                   ||
  12  GluAsnPro......GlyValThrGlnLeuAsn............ArgLe  22

2264  GAGCCTGGTAATCCACCTGCAGAACTCCCCGCATTCCAGTGCATGGAGCC  2215
      |::::              || |::::::::  :::::::|||||
  22  uAlaAlaHisProProPheAlaSerAspProMetGlyAlaProGlyThrP  39

2214  CTTCT.........GGCTCCCCTGTATAAGTCCAGACTGAAACC  2180
      ||::              ||| |||           ||
  39  roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly  55

2179  CCCTTGGAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGAGACGAGGAG  2130
      |||  :::|||||             |||||||
  56  Pro...GlnGlyLeu...ProGlySerPro............GlyAlaPr  66

2129  AGGGACGCGCCCAGCTGTGCAGCTACGCACCTCAGCA.......G  2086
      ||||||||||  ||  ::::::                ::::||
  66  oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT  83

2085  CACAGGGTGGCAGCAGGAGCCACCATTACTTTGGCAGCAA..........  2046
      ::  ||||||::::::::::                       |||||
  83  hrProGlyGlyAlaLysGlnAlaLysGlyLeuGlySerProGlyAlaPro  99

2045  ...........CAGAAACTGGCGCCGCAGCCCGGCCAGCCCATGGGG  2011
                 ::  ||||||  :::|||||::::||||||
 100  GlyThrProGlyGlnGlyLeuProGlySerProGlyAlaProGlyTh  116

2010  CTAACAGGAGGCGGGGAGCTGGGACCCAGTGAGGCAGCCCTCCACCCCAA  1961
      ::|||         ||  ||  |||:::||||
 116  rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro.  132

1960  TGTGTCTGGAAGTTTCTCACGCTGAGTATTGGCCAAGTCGCTCTTGTCAA  1911
                                             :::    |||
 133  ................................GlyProGln  135

1910  ATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGGCGTCGCGCA  1861
                                          ::::||||||
 136  GlyLeuPro.............................GlySerProGlyAlaProGl  145

1860  GACACCATATAGGCAGTGACAGACTGGCTGAGCTGACAATGGAGCCCAT  1811
      ||||||        ||::|||  :::|||  ::|||      ||||||
 145  yThrProGlyGlyAlaLysGlnAlaGlyAspValGly......SerPro.  159

1810  AAACAGGGATGGGCACCACCTGGGACAGCAGCAGGAGGCAC..........  1773
      ::|||               ::    |||     ::
 160  ...........GlyAlaProGlyThrProGlyProGlnGlyLeuProGly  172

1772  .........TATCCAGGATGGCGAGGTCCAGGCCAGATGCCCCGGC.....  1737
      ||||||                           ||||||
 173  SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr  189

1736  ......CCGGAACCACCCTGGCCTCGGTG.......GGCTCACCACCA  1701
      ||| |||::   |||:::   |||        |||||||||
 189  oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA  206

1700  CCACACGTACG.......................GAGACATCACAGGCAGAGGCCCG  1666
      ::                               :::::::::::  :::|||
```

```
551  ......SerProGlyAlaProGlyThrProGlyPro...........  560

458  TGCCCAGCACCATGGTCATGAACTTCCTCTACCCCCACTTCCAGCAGC  409
     :::   |||||||           ||:::||
561  ..............GlnGlyLeuProGlySerProGlyAl  569

561  ......................ProGlyThrProGlnG  577
569  a.................

408  AGAGGCGGCACATAGGTGATGCCTGGGCCAAACACACTCCAGGCCAAA  359
     |       |||::::||||  |||
569  a..............ProGlyThrProGlnG  577

358  GGTTAGGAGGTTGACCAGCAAGAGCTGGGCTTTCCGGTGCCCA......  315
     :::::|||  |||:::            ||
577  lyLeuProGlySerProGlyAlaProGlyThrProGlyAlaLysGln  593

314  GCAGGC......GGCTCACCCAGCCTCTGGACCATAGTGGGCCAGGCG  271
     ||||||      ||||||||  |||    ||||||    |
594  AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG  610

270  GG..TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAAACTGCTTCGTCTC  224
     || |||||  |||          |||||||||
610  lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu  626

223  GGCTCTGCTCCAGAAGCTGCGGGCCTCCTCCTCCTTGCTGCCGCCAACTGCC  174
     |||::::|||  ||||||:::|||||                 ||
627  ProGlySerProGlyAlaProGlyThrProGlyGly.................Pr  638

173  TAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAG  124
     |::::::  |||||||::          |||   |||
638  oGlnGlyLeuProGlySer..........ProGlyAlaProGlyThrP  651

123  CTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGACCACCTCAGTGGGGA  74
     ::                                    |||
651  roGly.............................GlyAlaLysGlnAlaGly  658

73   CACGTCTCATCA..........CTCAGATCCTGCCGA  46
     |||||:::||          ::: ||||||||||
659  AspValGlySerProGlyAlaMet.AspProGlyArg  670
```

seq_name: A_Geneseq_36:W49717

```
seq_documentation_block:
ID   W49717 standard; Protein; 682 AA.
AC   W49717;
DT   12-OCT-1998 (first entry)
DE   Protein polymer adhesive substrate PPAS1-C.
DE   Protein polymer adhesive substrate; PPAS1-C; sealant;
KW   wound healing; CLP 3.7; collagen; human.
OS   Synthetic.
OS   Homo sapiens.
PN   US5773577-A.
PD   30-JUN-1998.
PF   03-MAR-1994; 205518.
PR   02-MAR-1995; US-397633.
PR   03-MAR-1994; US-205518.
PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
DR   WPI; 98-387091/33.
PT   New recombinant protein polymers - containing naturally occurring
PT   repetitive units for crosslinking by enzymes, useful as medical
PT   adhesives and sealants, depots and matrices
PS   Example 4; Column 29; 70pp; English.
CC   Protein polymer adhesive substrate PPAS1-C comprises 10 repeats
CC   of a monomer (see W49716) consisting of human fibrin gamma chain
CC   peptide 93.4 (see W49709) inserted into a structural backbone of
CC   collagen repeat motifs (see W49711). The 59 kDa PPAS1-C polymer
CC   has been expressed in Escherichia coli HB101 transformants using
CC   plasmid pPT0322. Claimed recombinant protein polymers are capable
CC   of covalent crosslinking by enzymatic reaction to form products
CC   which set quickly and have good adhesive properties and high
CC   strength. CLP (collagen like protein) polymers are extremely soluble
CC   in water, allowing protein solutions of over 10 wt.% to be formed
CC   while maintaining good flow properties. CLP polymers have good
CC   adhesion to hydrophilic surfaces and therefore may adhere well to
```

```
206 laProGlyThrProGlyGlyAlaLysGlnAlaGlyLysAspValGlySerPro 222
1665 CAGGACCGCGGGTGGAGGTGGGAGGCAGGCCACTGCCTTCCAGCACCCACGTG 1616
       ::: |||  ||  |||||  ||   :::|||
223 GlyAlaProGlyThrProGlnGlyLeuProGlySerProGlyAl 239

1615 TCCATTAGGGAAGGGAGCTTCCAGGCTTAGGGCCTGGCAGGAAGCTGGTCA 1566
       ::||  |||  |||||  ||||||   :::||
239 aProGlyThrProGlnGlyLeu...ProGly.SerProGlyAla 254

1565 TCAGGCTGTCCTCACTGCTAGCACCTCCAGTGTCCCCTCGGTATTTGGGC 1516
       |||  |||  |||  |||::||
255 ProGlyLyThrProGly.............ProGlnGlyLeu...... 263

1515 AGGAACACCTGCTTCTCCGGTGGTAGAGGGAGGCCA.....GTGTGTA 1472
       |||:::||||||||||:::  |||  ::|||
264 ......ProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysG 278

1471 GGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAG 1422
       ::||||||:::||:||
278 lnAlaGlyAspValGly...... 283

1421 CTGTCACCACGGCCACACTGTGGACAGGCATGTGCACCGGCAGCCACA 1372
       |||||                                   ||||
284 ...SerProGlyAlaProGlyThrProGly........ProGl 294

1371 GGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTG 1322
       ::|||||||:::||:||                         :::|
294 nGly...LeuProGlySerPro.............GlyA 302

1321 CACCAGCCGGTCCATGACCAGAGAAGACACCAGGGAGATGGCCACTGCA 1272
       |||||:::  |||  ||:::||
302 laProGlyThrProGlnGlyLeuProGlnGlyLeuProGlySerProGlyAlaPro 318

1271 GGAACA...............................GCCCAGGCTG 1255
       ||||||
319 GlyThrProGlyProGlnGlyLeuProGlnGlyLeuProGlyGlySerProGlyAlaProGlyTh 335

1254 CCCA...TCCGAACGCCTTCATCATAGTGTCTCCGGCGTTCGGTGCCGG 1208
       ::||                                   |||
335 rProGlyGlyAlaLysAlaLysGlnAlaLysGlnAlaGlyLysAspValGlySerProGlyAlaProG 352

1207 CTCACGCTCTGGGCACGC.....CCTGGTACAGCCCCTCGCCCACGAAAT 1164
       ::||                ||||||:::|||
352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368

1163 CCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGC............ 1125
       |||                                ::||||:::|||
369 ProGlyProGlnGly.............LeuProGlySerProGlyAlaPr 381

1124 .TGCACAGCTCAGCCACGAGAGCGCGGCCAGGGTGCGGGGCCATGCGGGCA 1076
       |||  ||||||  :::||
381 oGlyThrProGlyProGlnGlyLeuProGlnGlyLeuProGlySerProGlyAlaProGlyT 398

1075 GCACAGCTGGTGCA.....GCCGGGGAAGCAGGCGCCCAGGTCC.... 1035
       ::  :::||||||              |||||||:::|||
398 hrProGlyProGlnGlyLeuProGlnGlyLeuProGlyAlaProGlyLys 414

1034 GGAAAGCCAAGCGGCCCGGCCATGGACAGCAGTGGGCGACAAGGAGGGG 985
       |||  |||:::  ::|||
415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431

984 GCCGACAGCCCTTCTGCTGGCTCCGGTGGGGGCCCA...GCCGTGCCTCCTC 938
       |||  |||:::  ||||||:::
431 rPro............GlyProGlnGlyLeuProGlyS 440

937 AGCCACCAGCAGTGTGGCTGCTACCCAGGTGAGGATGAGGGTGAGCA 888
       :::|||  ||:::||:||                      |||  ::|||
440 erPro.............GlyAlaProGlyThrProGly 447

887 GGGCCAAGAGAGGCCACTCCT.....CCTGGGGTGCCCAGGTAGGGGGCCAGG 844
       |||  ::||:||              ||||||||::||||||:::  ||||
448 GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProGl 463
```

```
843 GCACTGGTGTCCCAGTCAATGGCAGGCAGGA.....GGTAGCCCAGGCA 800
      |                              ||||||||||||:
463 y.............GlyAlaLysGlnAlaLysGlyAspValGlySerProGlyA 476

799 GCCCCAAGACTGATCATGAGGCATAGACAGAGTAGGCCTGGCGACAGT 750
      ::|||                                   :::|| |||||
476 laPro.............GlyThrProGlyProGln 483

749 GGT...CCGGGT...CCCGGAAGAGGTCAGAGAGCAGGAGGCTCCAGTGGA 706
      ||| |||||| :::|| ||||||||:::||
484 GlyLeuProGlySerProGlyAlaProGlyThrProGlyGlnGlyLe 500

705 GTGAAGCACACCTGGCCACAGAGTCCAGCAGCCCACGCCCAGGATGAG 656
      :::||||||||| :::||||||||                   |||
500 uProGlyThrProGlyGly.......AlaProGlyThrProGlyProGlnGlyL 515

655 CAGTGCCAGTCCAGGGCCTGGGATCCGGGCACAGCAGCCCTGCTAGCC 606
      |||||||                   ||||||:::        |||||
515 euProGlySerProGlyAla.......ProGlyThrProGly.....Gly 527

605 AGCCGGCCCTTGGGATGAGAAGAGGCTCAGCAGGATGCCCAAGGACAGT 556
      ::::::|||||     ||:::||||||:::|||
528 AlaLysGlnAlaGly...AspValGlySerProGlyAlaProGlyThrPr 543

555 G...CCCAGATGAAGGGCCGGCGCCCATAGCGTCCCAGTGCC 509
      |||:::|||
543 oGlyProGlnGlyLeuProGly............. 550

508 ACTGGGTGAGCCTAGGAGCGGGACACAGACCAGGCCACACTGGACCAA 459
      |||                                :::||||||
551 .......SerProGlyAlaProGlyThrProGlyPro....... 560

458 TGCCCAGCACCATGGTCATGAACTTCTCCTGACCCACTTCCAGCAGC 409
      |||||||                           |||::::|
561 ..............GlnGlyLeuProGlySerProGlyAl 569

408 AGAGGGCGCACATAGGTGATGCCTGCGGCCAAACACACCTCAGGCCAAA 359
      :::  :::|||||                         |||
569 a.......ProGlyThrProGlyProGlnG 577

358 GGTTAGCAGGTTGACCACGACAAGAGCTGGGCTTTCCGGTGCCGCA 315
      |||  :::|||  |||||||:::|||
577 lyLeuProGlySerProGlyAlaProGlyThrProGlyAlaProGlyAlaLysGln 593

314 GCCAGGC.....GGCTCACCCACCACCTCTGACCATAGTGGGCCAGGCCG 271
      |||||||             ||||||||:::||||::::||
594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610

270 GG...TAGGGCTCAGGGGCCGTTCAGGCACTCCAGGAACTGCTTCGTCTC 224
      ||  :::|||||  ||||||  :::||
610 lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu 626

223 GGCTCTGCTGCCAGAAGCTGCGGCCTCCTCCTCCTGCGCCCAACTGCC 174
      :::|||||  ||::::::|||
627 ProGlySerProGlyAlaProGlyThrProGly.............Pr 638

173 TAGGAATCAGCCAGGGCGCCCATTTCTGCCAGCCCCTTTGGTGCCGGTCCAG 124
      |||:::||||  :::||||||                       |||  ::|||
638 oGlnGlyLeuProGlySer.............ProGlyAlaProGlyThrP 651

123 CTTCTCAGCCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTGGGA 74
      :::                                   ||||||  |||||
651 roGly.............GlyAlaLysGlnAlaProGlyThrProGlyGly 658

73 CACGTCTCATCA........CTCAGATCCTGGCCGA 46
     |||||:::|||              :::  ||||||||||||
659 AspValGlySerProGlyAlaMet.AspProGlyArg 670
```

seq_name: A_Geneseq_36:R28916

seq_documentation_block:
ID R28916 standard; Protein; 1196 AA.

```
AC   R28916;
DT   24-MAR-1993  (first entry)
DE   Type III procollagen (prior art).
KW   Mutation; pro-alpha1(III); primer; PCR.
OS   Homo sapiens.
PN   WO9219754-A.
PD   12-NOV-1992.
PF   08-MAY-1992; U03866.
PR   08-MAY-1991; US-696607.
PA   (UTJE-) UNIV JEFFERSON THOMAS.
PI   Kuivaniemi SH, Prockop DJ, Tromp GC;
DR   WPI; 92-398878/48.
DR   N-PSDB; Q30849.
PT   Kit for detecting genetic pre-disposition for vascular aneurysms
PT   - contains primer to amplify portions of Type III procollagen DNA
PT   and detects mutation in standard procollagen DNA
PS   Disclosure; Fig 1A-F; 44pp; English.
CC   Example 1 describes the determination of the presence of a mutation
CC   in the pro-alpha1(III) gene.  Primers used in PCR are given in
CC   Q30834-48.
SQ   Sequence  1196 AA;
```

```
alignment_scores:
     Quality:   368.00        Length:   1018
     Ratio:     0.860         Gaps:       67
Percent Similarity: 42.043    Percent Identity: 26.031
```

```
alignment_block:
US-09-030-606-110/rev x R28916

Align seg 1/1  to: R28916  from: 1 to: 1196

2591 CTGTGTCTCTGTGATGGC...........................        2574
     ::||||||||||||||
52   IleCysLeuCysAspSerGlySerLeuLeuSerAspAspIleIleCysAs  68

2573 .AACAGAAGGACCACAGGCCACATCCTGATAAAAGGTAAGAGGGGGGTG   2525
     :::::::  ::: |||::|||:||
68   pAspGlnGluLeuAspCysProAsnProGlu..................  78

2524 GATCAGCAAAAAGACCAGTGCTGT.....................       2502
     ||:|||||||
79   ..IleProPheArgGluCysCysAlaValCysProGlnProProThrVal  94

2501 .........................GGGCTGAGGGGA...CC         2488
                              |||  :::||| |||
95   ProThrProProAsnGlyGlnGlyProGlnGlyProLysGlyAspPr     111

2487 TGGTTCTTGTGTGTGCCCCTCAGG.....................       2463
     ||||    ::|||||
111  oGlyProProGlyIleProGlyArgAsnAlaAspProGlyIleProGlyG  128

2463 .................................                   2463
128  lnProGlySerProGlyProProGlyIleCysGluSerCys           144

2462 ...ACTCTTCCCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGT  2416
     |||  :::  ||:::||||||
145  ProThrGlyProGlnAsnTyrSerArgGlnTyrAspSer..........   157

2415 GTTTCATCCTAGAAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAG 2366
     ::|||||||||||||||  ::: |||
158  .TyrHisValLysSerGlyLeuAlaValGlyLeuAlaGlyTyrProG     174

2365 GTTAAGGGGCTTAGAGATGGGAAAACCAGGTGACTGAGTTTATTCAGCT.. 2318
     ||   :::|||||| |||||  ||||||
174  lyProAlaGly.ProProGlyProProGlyProAlathrSerAlaHi     190

2317 .CCCAAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAA  2269
     |||   |||||
190  sProGlySerProGlyTyrGln......................        200
```

```
2268 CCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATTCCAGTGCCATGG 2219
     ::|||||      ::
201  ..........GlyProProArgGlnProGlyGlnAlaGlyProSer       212

2218 AGCCCTTCTGGCCTCCCCGTATAAGTCCAGACTGAAACCCCCTG.....   2174
     ::|||  |||     :::::  |||
213  GlyProProGlyProProGlyProIleArgProSerArgProAlaGlyLy  229

2173 .GAAGGCTCCAGTCAGGCAGCCCTAGAGACTGGGAGAGACTGGGAGAGGGA 2125
     ::|||    :::|||||   |||      |||||
229  sAspGly...GluSerGlyArgProGlyArgProGlyGlu...ArgGlyL  244

2124 CGCCCCAGCCCCAGCTGCAGCTAGCCACCTCAGCAGCACAGGGTGC      2075
     |||  |||||:|:::::::     :::
244  euProGlyProProGlyIleLysGlyProAlaGlyIleProGlyPhePro  260

2074 AGCAGAGGCCACAT......TACTTTGCAGCAACAGAAACTGGCGGCC    2031
     :::  ::: |||     |||  |||   ||| :::|:::
261  GlyMetLysGlyHisArgGlyPheAspGlyArgAsnGlyGlu.AsnGlyG  277

2030 AGCCCGGCAGCCCCATGGGCAACAGAGACGGGGACTGGGACCCAGTG     1981
     :: |||:::|||  :::::|||    :::
277  luThrGlyAlaPro...GlyLeuAsnGlyGluAsnGlyLeuGlnGlyGlu  292

1980 AGGCAGGCCCTCACCCCAATGTGCTGGAAGTTTCTCACGCTGAGTATTT   1931
     ::: ||| |||||:::|||
293  AsnProAla.ProArgProMetAla.........................  300

1930 GGCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGA  1881
     ||| ||:::                          ::|:::  |||
301  .ProArgArgAla....................  ProGlyGluArgArg  309

1880 CCAGACCCAGGCCTGCGGCAGCACCATATAGGCAGTGACAGACTGGCTG   1831
     ||  ||||||      :::|||    :::
310  Pro...ProGlyLeuProGlyAlaProGlyAlaArgGlyAsnAspGlyAl  325

1830 AGCTGGA.....CAATGGAGCCCATAAACAGGG.................  1803
     ||| :::    |||       :::
325  aArgGlySerAspArgGlnProProProGlyProProGlyThrAlaA     342

1802 ................................ATGGGGCCACCTGGGACA  1785
                                      :::|||||||||||:::
342  laPheProGlySerProGlyAlaLysGlyGluValGlyProProGlySer  358

1784 GCAGGAAGGCAC..TATCCAGGATGGCGAGGT..................  1755
     ||: :::                 ||||||
359  ProGlySerAsnGlyAlaProGlyGlnArgGlyGluProGlyProGlnAl  375

1754 ....CCAGGCAGATGCCCCGGCGAACCACCCTGGCCTCGGTGGGCTC     1709
     :::|||||| ||| ||| |||| ||||  ||||
375  aHisProGlyAlaGlnProProProAlaPro.ProGlyIleAsn.GlySe  391

1708 ACCCACCACCACAGTACGGAGACATCACAGGCAGAGGCCCGCAGAGCG    1659
     |||: |||||||||||||||:::
391  rPro...GlyGlyLysGlyGluMetGlyProAlaGlyIleProGlyAlaP  407

1658 CGGGTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCACCACGTGTCCATTA 1609
     |||:||| |||||:::|||
407  roGlyLeuMetGlyAlaArgGly...ProProGlyProAlaGly......  420

1608 GGGAAGGGAGCTCCAGGCTTAGGGCCCTGGCCAGGAAGCTGGTCATCAGGCT 1559
     :::||||||||||||||| ||| :::
421  AlaAsnGlyAlaProGlyLeuArgProGlyAlaArg..............  432

1558 GTCCTCACTGCTAGCACCTCCAGTGTCCCCTGGTATTGGGCA......    1515
     ||||||      |||
433  ..........GluProArgLysAsn.AlaAlaLys                 440

1514 GGAACACCTGCTTCTCCCGGTGTGAGAGGGAGGCCAGTGTGTAGGGCAGG  1465
     :::||||| |||||
441  GlyGluProAlaProProGlyGluArgGly.......GluAlaGl       453

1464 ATCTGCAGGCTGAGAAGGTGAACCCGGTGAGGGCGGCCTGAAGCTGTCAC  1415
```

453 ylIeProGlyVal..457

1414 CACGGCCACACTGTGGGACAGGCATGTGCACCGGCACCGGCAGGGAAAG 1365
     |||||||:::|||
458 ....ProGlyProLysGlyGlu 463

1364 CTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGTCGACCAGC 1315
     ::::||||| ::|||:||||
464 ....AspAlaLysAspGlySerProGl 471

1314 CGGTCCATGACCAGAGAAGAAGAGATGGCGCACTGCAGGAAACAG 1265
     : |||||| ||||||
471 yGluProGlyAlaAsnGlyLeuProGlyAlaAlaGlyGluArgGly...A 487

1264 CCCCAGGCTGCCCATCCGAACGCCTTCATCATAGTGTCTCCGGGCCTCGG 1215
     |||||||| :::||||: |||
487 laProGlyPheArgGlyProAlaGlyAsnGlyIleProGlyGluLys 503

1214 TGCCCGGCTCAGCTCTGGCACGCCCTGGTACAGCCCCTCGCCACGAAA 1165
     |||||| |:: |||||| ||||||||:::
504 GlyProAlaGlyGluArgGlyAlaProGlyAla...ArgProArgGl 519

1164 TCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCC......AGCTGCA 1121
     |||||||:::||:
519 yAlaAlaGlyGluProGlyArgAspGlyValProGlyGlyProGlyMetA 536

1120 CAGCTCAGCCACGAAGAGCGCGGCCAGGGTGCGGGGGCATGCGGCAGCACA 1071
     :::::||| ||| ||||| |||
536 rgGlyMetProGlySerProGly...GlyProGlySerAspGlyLysPro 551

1070 GCT.....GGTGCAGCCGGGGAAGCAGGGCGCCCAGGTTCCGGAAGCC 1027
     ::: |||||:: ||||||||||||||
552 GlyProProGlySerGlnGlyGluSerGlyArgProGlyArgProGlyPr 568

1026 AAGCGGGCCC.................................1017
     ||||| |||||||
568 oSerGlyProArgGlyIleProGlyValMetGlyPheProGlyProLysG 585

1016 ......GGCATGGACAGCAGTGGGGCGGCCACAAGGAGGGGGCCGACAGCC 975
     ||| :::::::::: ||| ||||||
585 lyAsnAspGlyLysAsnGlyLeuArgGlyArgGlyPro......599

974 CTTCTGCTCGGTCGGGGAAGCAGGGCGCTGCCTCCTCAGCCACCAGCAGT 925
     |||||||::: |||
600 ....GlyGlyProGly...ProGlnGlyProPro.....608

924 GTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCA 875
     |||:::|||||:: ||||||
609 ....GlyLysAsnGlyLysGluThrGlyThrProGlnGly. 618

874 CTCCTCCTGGTGCCCAGTAGGGGG.......................849
     |||||||||| |||
648 yProProGly..................ProProGly 634

810 TAGCCCCAGGCAGCCCCAAGACTGATCATGATGAAGGCATAGAAGCAGAGTAGGC 761
     ::::::::::::
648 yProProGly.........................GluAsnGlyLysP 656

760 CTGGCGACAGTGTCCG.....GGTCCCGGA 735
     |||||| |||||||||
656 roGlyGluProGlyProLysGlyProGlyAspAlaGlyAlaProGlyGly 672

734 AGAGGTCAGAGACGAGAGGGCCTCCAGTGGAGTGAAGCAACCTGCCACAG 685
     |||||| ||| ::: |||||
673 GlyLysGlyAspAlaGlyAlaProGlyGlyArgGlyProGlyProGlyLeuAl 689

684 AAGTCCAGGAGCAGCCCACGCCCAGGATGAGCAGTGCCAGCTCCA.......642
     ::: ||| ::: |||:::|||

689 aGlyAlaProGlyLeuArgGlyGlyAlaGlyProProGlyProGluGlyG 706

641 .......GGGGCTGGGATCCGGCACCACCACCGCCCTGCTCAGCCAGCCGG 600
     ::::||
706 lyLysGlyAlaAlaGlyProProGlyProPro..................716

599 CCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCA......AGGACAGT 556
     ||| ::::|||||
717 ....GlyAlaAlaGlyThrProGlyLeuGlnGlyMe 727

555 GCCCAGATGAAGGGCCGGCGGCCCATAGCGTCCACGCCCAGTGGTCACT 506
     :||| |||:::|||| ||| ::::|||
727 tProGlyGluArgGlyGlyLeuGly...SerProGlyProLysGly.... 741

505 GGCTGAGCCTAGGAGCCGGGACACAGCCCAGGCCACTGGACCAATGC 456
     ||||||||:::|||||:: |||
742 ....AspLysGlyGlyProGlyGlyProGlyAlaAspGlyVal 754

455 CCA........GCACCATGTCATGAACTTCTCCTCTA...CCCCCACT 418
     ||| |||||||||||||
755 ProGlyLysAspGlyProArgGlyProThrGlyProIleGlyProProGl 771

417 TCCAGCAGCAGAGGCGGCACATAGGTGATGCGTCGCGGCCAAACACACCTC 368
     ||||||:::::::
771 yProAlaGlyGln...................P 776

367 CAGGCCAAAGG...TTAGCAGGTTGACCAGCAAGAGCTGGCTTTCCGGT 321
     ||||:::||| |||:::: ||||||:::||
776 roGlyAspLysGlyGlyGlyAlaProGlyLeuProGlyIleAlaGly 792

320 GCCGCAGCAGGCGGCTCACCCACAGCCTCTGGACCATAGTGGGCCAGGCG 271
     ||||||||| ::||||||||||||||
793 ProArg......GlySerProGlyGluArgGly.GluSerGlyProProG 807

270 GG..........TAGGGCTCAGGGGGCCGTTCAGGCACTCCA.......239
     || |||||:::||| |||
807 lyProProAlaPheProGlyAlaProGlyGlnAsnGlyGlnProGlyGly 823

238 ......GAACTGCTTCGTCTCGGC....................TCTGCTCCAGAAGC 207
     ||||| :::::|||
824 LysGlyGluArgGlyAlaLeuGlyLeuLysGlyGluGlyProGl 840

206 TGGCGGCCTCTCCTCCTTGCCTGCCGCGCCAACTGCCTAGGAATCAGCCAGGCCG 157
     ::::||||||||||:::
840 yValAlaGlyPro......................844

156 CCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCTCAGCCCCATGCTC 107
     ||| |||:::||||||||:::
845 ....ProGlyGlySerGlyProAlaGlyProPro... 854

106 AACACCTGCTGCTGTGGGGGACACCTCAGTGGGGACACGTCTCATCACTCAG 57
     :::||||||| ||| ||||||
855 ....GlyProGln.GlyValLysGlyGlyArgGlyS 865

56 ATCCT...GGCCAGGCGCCGGCGTCACCCGGAGCC 22
    ::|||| || |||||| |||||
865 erProGlyProGlyAlaAlaGlyPhePROGlyAla 877

Wed Sep 29 14:26:52 1999

PA   (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PI   Cappello J;
DR   WPI; 95-320413/41.
PT   Protein polymers comprising repeating units and sequences - capable
PT   of enzyme-catalysed covalent bond formation useful as a
PT   biocompatible material for wound closure and tissue repair
PS   Example 4; Page 43; 138pp; English.
CC   The amino acid sequence of the protein polymeric adhesion substrate
CC   (PPAS) 1-A.   The protein comprises 10 repeats of the PPAS1-A monomeric
CC   repeat (R80323) which consists of the CLP 3.7 gene sequence linked to the
CC   human fibrin cross-linking substrate sequence (see Q98704) corresponding
CC   to the wild type 17 C-terminal amino acids from human fibrin. The
CC   polymers can be used in biological systems where in situ formation of a
CC   biocompatible material with structural integrity is required e.g. as
CC   medical adhesives and sealants or for wound closure or tissue repair.
SQ   Sequence   762 AA;

alignment_scores:
    Quality: 365.00   Length: 751
    Ratio: 1.058   Gaps: 52
Percent Similarity: 45.939   Percent Identity: 30.093

alignment_block:
US-09-030-606-110 x R80324   ..

Align seg 1/1 to: R80324   from: 1 to: 762

```
 84 GGTGCCCCACCAGCAGGTGTG.......AGCATGGGCTGAGAAGCTGG 127
    |||::|||  |||              |||     ||| |||
 52 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh  68

128 ACCGGCACCAAAGGGC......TGGCAGAAATGGGCGCCTGGC....... 164
    |||:||||:||:||:||       ::  ||||||||
 68 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG  85

165 ............TGATTCCTAGGCAGTTGGCGGCAGCAGGAGGAGAG 200
                ||||||||  ||| ::|||::||:|||
 85 lyGluGlyGlnGlnHisHisLeuGlyGly...AlaLysGlnAlaGlyAsp 100

201 GCCGCAGCGTTCTGAGCAGAGCCAGCAGAGCAAGCAGTTCTGGAGTGCCTGA 250
    ::|:::||  ||||||         ::  ::||||||||::|||
101 ValGlySerProGlyAlaProGlyThrProGlnGlyLeuProGl 117

251 ACGGCCCCCTGAGCCCTACCCGCCTGGCCCACTATGGT........... 288
    ||||||||||  ||||||||||  |||
117 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 133

289 CCAGAGGCTGTGGGTGAGCCG...CCTGCTGCGGCCACCGGAAAGCCCAGC 335
    ||| ||| |||  ||| ::|   ||| |||::||||:
134 ProGlyAlaProGlyThrProGlnGlyLeuProGlySerProGl 150

336 TCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGCCCGCA 385
    ||| ||:::|||          ::||||::|||||:
150 yAlaProGlyThrProGlyGlu.............GlyGlnG 160

386 GGCATCACCTATGTGCCGCCCTCTGCTGGTGGAAGTGGGGTAGAGGAGAA 435
    :||||||||  ||||||:::|||:||||||
160 lnHisHisLeuGlyGlyLysGlyAlaAlaLysGlyAspValGly........ 173

436 GTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTGCCTGGTCTGTG 485
    |||          ||  ::|:|||||
174 .......................SerProGlyAlaProGly...Th 180

486 TCCCGCCTTCATCTGGCCACTGTCCTTGGCATCCTGCTGAGCCT... 535
    ||:::||||:::                ||| :||::  ||
180 rProGlyProGln......................... 184

536 CGCCGGCCCTTCATCTGGCACTGTCCTTGGCATCCTGCTGAGCCT... 582
    ||: |||::||  |||::|
185 ..........GlyLeuProGlySerProGlyAlaProGly 194

583 CTTTCTCATCCCAAGGGCCCGGCTGGCTAGCAGGGGCTGCT.....GTGCC 626
```
```
195 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211
    |||:||| ||:                            :||::|||
627 CGGATCCCAG.......GCCCCTGGAGCT...GGCACTGCTCATC 661
    |||::|||   ||||||||      |||||     |||||
211 OGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr......P 226

662 CTGGGCGTGGG...........GCTGCTGGACTTCTGTGGCCAGGTGTG 699
    ||||::||||                 ||  :|||
226 roGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAsp 242

700 CTTCACTCCACTGGAGGCCCTGCTCTCTGACCTCTTCCGGGACCC...... 744
    :::  |||::||:::|||      ||||||||
243 ValGlySerProGlyAlaProGly.........ThrProGlyProGlnGl 256

745 .................GGACCTGTCGCCCAGGCCTACTCTGTCTAT 775
                     ||||::||||||
256 yLeuProGlySerProGlyAlaProGlyThrProGly........... 268

776 GCCTTCATGATCAGTCTTGGGGGCTGCCTGGCTACCTCCTGCCTGCCAT 825
    |||||||||||| |||      ||| |||
269 ............ProGlnGlyLeuProGlySerProGlyAla....... 278

826 TGACTGGGACACCAGTGCCCTGGCCCC......CTACCTGGCACCCAGG 869
    ||||||           :::|||
279 ..........ProGlyThrProGlnGlyLeuProGlySerPro.. 291

870 AGGAGTGCCTCTTGCCTGCTCACCCTCATCTTCCTCACCTGCCTAGCA 919
    ||  ::||||     |||:||   ||:||:||  :::
292 .GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyA 308

920 GCCACACTGCTGTGGCTGAGGAGGCAGCCTGGGCCCCACCGAGCCAGC 969
    ||||||           |||:::|||   :::|||
308 laLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324

970 AGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGCTGTCCATGCCGGG 1019
    |||  :||||          |||||||| |||||||||
325 ProGlnGlyLeuProGlySerPro......GlyAlaProGly 339

1020 CCCGGCTTGGCTTCCG...GAACCTGGGCGCCCTGCTCCCCGGCTGCGAC 1066
     |||| |||:::|||||   ||| :||:||||||      ::::|
339 yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 355

1067 CAGCTGTCTGCGCCATGCCCCGCACCCTGCGCCGGCTCTTCGTGGCTGA 1116
     ||    |||::|            |||||||:::|||::||||||
355 roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 370

1117 GCTGTGCAG.........CTGGATGGCACTCATGACCTTCACGCTGTT..... 1155
     |||:|||        |||: ::  ::  |:||||
371 GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 387

1156 .........TTACACGGATTCGTGGGCGAGGGGCTGTACCAGGGCGTGCCC 1198
              |||     ||| :::::|||||::|||
387 rProGlyAlaProGlyThrProGlnGlyLeuProGlySerProG 404

1199 AGAGCTGAGCCGCGGCACCGAGCCCGGAGACACTGATGAAGGCGTTCG 1248
     |||||| ||:|||   |||     :::|||
404 lyAla....ProGlyThrProGlyProGlnGlyLeu........... 414

1249 GATGGCAGCCTGGGGCT.......GTTCCTGCAGTGCCATCTCCCTGG 1292
     |||:::||||||||||        ::::: || ||||||||||
415 ProGlySerProGlyAlaProGlyThrProGlnGlyLeuProGl 431

1293 TCTTCTCTCGTGCTCATGGACCGGCTGGTGCAGCGATTCGGCAC...... 1335
     |::::::|||  |||      ||  ||||||:::||
431 lySerProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuG 448

1336 ..TCGAGCAGTCTATTGGCCAGTGTGGCAGCTT....TCCCTGTGGCTG 1379
       ::||::||:|  |||||     ||||||    :::|||||||
448 lyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464

1380 CCCGGTGCCACATGCCTGTCCC.....................ACAGTGTG 1408
     ||| |||::||   |||
```

```
 465  ProGlyProGlyLeuProGlySerProGlyAlaProGlyThrProGl  481

1409  GCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTCACCTTCTCAGCCCT  1458
       |||       |||    |||||||||::|||:::  :::|||
 481  yPro.....GlnGlyLeuProGlySerProGlyAlaProGly.ThrPro  495

1459  GCAGATCCTGCTCCACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGG  1508
       ::::           |||||||||     ||||||||    |
 496  GlyPro.........GlnGlyLeuProGlySerProGlyAlaProGl  508

1509  TGTTCCT..................................GCCCAAATACCGA  1528
       |::::|||                                 ||| |||
 508  yThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAla.LysGlnAla  524

1529  GGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCT  1578
       ||||||::::|||::  :::::    |||           :::   ||
 525  GlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe  541

1579  GCCAGGCCCT.....AAGCCTGGAGCTCCCTTCCCTAAATGGACACGTGG  1622
       ||||||           |||||||||:::|||  ||||:::||| |
 541  uProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG  558

1623  GTGCTGGAGGCAGTGGCCTGCCTCCCACCTCCACCCCGGCTCTGCGGGGCC  1672
       ||::: |||:::    |||  ||| |||||  :::||| |||::::||
 558  lySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySer  574

1672  ........................................  1672
                                          
 575  ProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGl  591

1673  ....TCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGGTGAGCCCCA  1715
       :::|||   |||                       |||||  |||||
 591  yAlaLysGlnAlaGlyAsp...............ValGly.SerPro  601

1716  CCGAGGCCAGGGTGGTTCCGGGCCGGGGCATCTGCCTGA.........  1755
       ||| |||   |||  |||      ||||||||
 602  GlyAlaPro.GlyThrProGlyProGlnGlyLeuProGlySerProGlyA  618

1756  ..CCTCGCCATCCTGA...TAGTGCCTTCCTGCTGTCCCAGGTGGCCCC  1800
       ||  |||||       |||||||    |||||||:::::|||
 618  laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro  634

1801  ATCCCTGTTTATGGGCTCCATTGTCCAGCTGACCAGTCTGTCACTGCCT  1850
               |||::::|||
 635  GlyThrProGlyProGlnGlyLeuProGlySerPro..............  646

1851  ATATGGTGTCTGCCCGGCCCTGGGTCTGGGTCGC......CATTTACTTT  1894
       ||:::     |||||::::|||:::        )      ||||||  )
 647  ....GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeu....  660

1895  GCTACACAGGTAGTATTGACAAGAGCGACTTGGCCAAATACTCAGCCGTA  1944
       ::::|:::         |||||||::||||::||
 661  ........GlyGly.....AlaLysGlnAlaGlyAspValGlySerPro  672

1945  GAAAACTTCCAGCACACCATTGGGGGTGGAGGGGCCTCACTGGGTCCCAGC  1994
       ||| |||||||||    |::|||||||
 673  GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl  689

1995  TCCCCGCTCCTGTTAGCCCCATGGGGCTGCCGGGGCTGGCCG  2035
       | |::::    ::::||||::|||||||||||  ||
 689  a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro  702
```

This Page Blank (uspto)

OM of: US-09-030-606-110 to: Issued_Patents_AA:*   out_format : pfs

Date: Sep 25, 1999 4:00 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171616_29813/app_query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-110
Query length: 3410
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

score_list:

Sequence	Std Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-120	-	395.00	518.70	..	1.1e-22	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-120	-	395.00	518.70	-	1.1e-22	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-84	-	393.00	515.96	-	1.5e-22	761
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-114	-	391.00	513.20	-	2.2e-22	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-26	-	391.00	513.20	-	2.2e-22	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-126	-	370.50	486.19	-	7.8e-21	682
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-36	-	370.50	486.19	-	7.8e-21	682
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-84	+	365.00	477.48	-	2.1e-20	761
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-114	+	365.00	477.47	-	2.1e-20	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-26	+	365.00	477.47	-	2.1e-20	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-120	+	364.00	476.09	-	2.5e-20	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-31	+	364.00	476.09	-	2.5e-20	762
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-62	-	360.00	467.08	-	5.8e-20	1064
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-132	-	357.00	465.58	-	9.0e-20	829
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-53	-	357.00	465.58	-	9.0e-20	829
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-175-155-68	-	356.50	464.80	-	9.9e-20	837
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-477-509B-103	-	356.50	464.80	-	9.9e-20	837
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-101	-	356.50	464.80	-	9.9e-20	837
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-75	-	356.50	464.07	-	1.0e-19	897
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-50	-	356.50	464.07	-	1.0e-19	897
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-53	+	353.50	458.13	-	1.8e-19	1065
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-73	-	344.50	451.24	-	7.4e-19	633
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-62	+	335.00	432.72	-	4.7e-18	1064
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-175-155-68	+	333.00	437.59	-	5.2e-18	516
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-356-340-4	+	333.00	437.59	-	5.2e-18	516
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-132	+	331.50	430.54	-	8.1e-18	829
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-53	+	331.50	430.54	-	8.1e-18	829
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-73	+	331.00	431.90	-	8.2e-18	682
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-36	+	331.00	431.90	-	8.2e-18	682
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-356-340-2	+	327.50	429.85	-	1.4e-17	525
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-53	-	326.00	423.66	-	2.1e-17	777
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-53	+	324.50	421.60	-	2.7e-17	777
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-73	+	323.00	421.69	-	3.3e-17	633
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-175-155-68	+	323.00	418.75	-	3.6e-17	837
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-477-509B-103	+	323.00	418.75	-	3.6e-17	837
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-101	+	323.00	418.75	-	3.6e-17	837
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-75	+	323.00	418.75	-	3.6e-17	837
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-50	+	323.00	418.02	-	3.7e-17	897
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-52	-	319.00	417.47	-	6.4e-17	561
/cgn2_6/ptodata/2/iaa/PCTUS95-02251-12	320.00	408.90	-	7.4e-17	144
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-02251-12	-	316.00	406.59	-	1.4e-16	1065
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-52	+	313.50	409.91	-	1.7e-16	561
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-494-168-2	+	299.50	379.03	-	2.9e-15	1694

/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-609-716-65 - 294.00 386.46 - 4.7e-15 408
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-972-032-82 - 295.50 378.30 - 5.0e-15 1077
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-95 - 295.50 378.30 - 5.0e-15 1077

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-120

seq_documentation_block:
; Sequence 120, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
;   APPLICANT: CAPPELLO, Joseph
;   APPLICANT: FERRARI, Franco A.
;   TITLE OF INVENTION: High Molecular Weight Collagen-Like
;   TITLE OF INVENTION: Protein Polymers
;   NUMBER OF SEQUENCES: 135
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/642,255
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: ROWLAND, Bertram I.
;   REGISTRATION NUMBER: 20,015
;   REFERENCE/DOCKET NUMBER: A55556-3/BIR
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 494-8700
;   TELEFAX: (415) 494-8771
;   TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 120:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 762 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-642-255-120

alignment_scores:
          Quality: 395.00   Length: 854
          Ratio: 1.082   Gaps: 59
Percent Similarity: 42.740   Percent Identity: 30.211

alignment_block:
US-09-030-606-110/rev x US-08-642-255-120

Align seg 1/1 to: US-08-642-255-120 from: 1 to: 762

2347 GGGAAACCAGGTGACTGAGTTTATTCAGGTCCCAAAAACCTTCTCTAGG 2298
     ||:::::|||||| ::|||:|||:::||::::
  76 GlySerProGly..........AlaProGlyThrProGlyGluGl 87

2297 TGTGTCT...CAACTAGGAGGCTAGCCTGTGTAACCCTGAGCCTGGGTAATC 2251
     | ::||||||||||| :::||||||||:|||
  87 yGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerP 104

2250 CACCTGCA.........GAGTCCCGCCATTCCAGTGCATGGAGCCCTTCT 2210
     || ||| ::|||::|||::::: |||||
 104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119

2209 GGCCTCCCGTGTATAAGTCAGACTGAAACCCCCTTGGAGGCCTCCAGTC 2160
     ||| ||| :::|||||

Wed Sep 29 14:26:53 1999

```
1311 TCCATGACCAGAGAGAGACCAGGGATGGCGCACTGCAGGAACA....  1266
      |||  |||::|||                     ::  ||||||
 337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG  354
1265 ...................................GCCCCAGGCTGCCATCCGAA  1245
                                        ||||||||| |||::||
 354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu  370
1244 CGGCTTCATCATAGTGTCTCCGGCCTCGGTGCCCGGCTCAGCTCTGGGC  1195
      :::::  |||  ::|||||| |||::|||         ::|::|:|
 371 GlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe  387
1194 A......CGCCCTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG  1151
      :                    |||  |||::|:::
 387 rProGlyAlaProGlyThr...ProGlyProGlnGly..............  398
1150 CGTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGAGCC  1101
      ||||:::||:|||           ::||||||||
 399 .......LeuProGlySerProGlyAlaProGlyThrProGly  410
1100 GGGCCAGGGTGCGGGGCATGCGGCAGCACAGCTGGTGCCGGGGAAGC  1051
      |||          ::  |||::  |||::  |||
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl  427
1050 AGGGGCCCCAGGTCC...GGAAAGCCAAGCGGGCCCCGC..........  1014
      |||  |||||||  |||:::  |||||
 427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG  444
1013 ...................................ATGACAGCAGTGGG  999
                                        :::  |||  |||
 444 lnHisHisLeuGlyGlyAlaAlaArgGlnAlaGlyAspValGlySerProGly  460
 998 GCGACA...AGGAGGGGCCGACAGCCCTTCTGCTGGCTCGGTGGGGCC  952
      |||      |||||||  :::|||                     ::|||
 461 AlaProGlyThrProGlyProGlnGlyLeu.........ProGlySerPr  474
 951 AGCGGCTGCCTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAGTGAGAA  902
      |:::  |||  |||  |||||||||||||||  ||| ||
 474 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA  491
 901 GATGAGGGTGAGCAGGCCAAAGAGGCACTCCT..............CCT  867
      :::  |||  |||              |||  |||
 491 laProGlyThrProGlnGlyLeuProGlySerProGlyAlaPro......  507
 866 GGGTGCCCAGGTAGGGGCCCAGGGACTGGTGTCCCAGTCAATGGCAGGC  817
      ||| |||:::|||  |||                     :::  |||
 508 GlyThrProGlyGluGlyGlnHisHisLeuGlyGlyAlaArgGlnAl  524
 816 AGGA......GGTAGCCCAGGCAGCGCCCCAAGACTGATCATGAAGGCATA  773
      ||||       |.||||||||||||:::|||
 524 aGlyAspValGlySerProGlyAlaPro...................  533
 772 GACAGAGTAGGCCTGGCGACAGTGGT...CCGGGT...CCCGGAAGAGGT  729
      :::  |||||||  ::  |||||||  |||||
 534 .....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro  548
 728 CAGAGAGCAGGGCCTCCAGGTGGATGGAGTGAAGCACACCTGGCCACAGAGTCC  679
      |||||||                                  ::||:||||||||
 549 GlyThrProGlyProGlnGlyLeuProGlySerProGly......AlaPr  563
 678 AGCAGCCCCACGCCCACGGATGAGCAGTGCCAGCTCCAGGGGCCTGGATC  629
      ||||||||||||||                      :::||||||||||
 563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.......P  578
 628 CGGGCACAGCGCCCTGCTAGCCAGCCCGGCCCTTGGGATGAGAAAGA...  582
      ||||||| |||           :::||| ||||||           |||
 578 roGlyThrProGlyGluGlyGlnHisHisLeuGlyGlyAlaArgGln  594
 581 ..........GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATG  547
      |||             |||  |||||||||       |||  |||
 595 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl  611
```

```
 120 GlyAlaPro.........GlyThrProGlyPro...GlnGlyLeu...Pr  131
2159 AGGCAGCCCTAGAGACTGGGGAGAGGAGAGGACGGCCCAGCCCCCAG  2110
      ||||||||  |||                     |||   :
 131 oGlySerPro.........GlyAlaProGlyThrProGlyProGlnG  144
2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT.....GGCAGCAGAGAG  2066
      :::::  :::|||   |||                  ||| |||
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln  159
2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGCGGCCAGCCCGGCCAGCCCCA  2016
      ||||||||||:||
 160 GlnHisHisLeuGly..................................  164
2015 TGGGGCTAACAGGAGCGGGGGAGCTGGGACCCAGTGAGGCAGCCCTCCAC  1966
      ||||||:::|||::|||                ||| |||::|:
 165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh  180
1965 CCCAATGTGCTGGAAGTTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTT  1916
      ||||                              |||     :::
 180 rPro...............................GlyP  183
1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGGCCTG  1866
      |||  ||||||          :::||||||
 183 roGlnGlyLeuPro...................GlySerProGlyAla  192
1865 CGGCAGACACCATATAGGCAGTGACAGACTGCGTGAGCTGACAATGGAG  1816
      ||||||                            |||
 193 ProGlyThrProGlyGln..............................  199
1815 CCCATAAACAGGATGGGGCCACCTGGGACAGCAGCAGGAAGGCCACTATCCAG  1766
      |||  |||||||::  |||                        |||
 200 ...............GlyLeuProGlySerProGlyAla......ProG  209
1765 GATGGGCCAGGTCCAGCAGATGCCCGGC..........CCGGAACCA  1728
      ||  |||||| |||||||                   ||| :::
 209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh  225
1727 CCCTGGCCTCCGGTGGGCTCACCCACCACCACCACGTACGGAGACATCACAG  1678
      ||||||  |||          ::|||||||  ||| |||  ||
 225 rProGlyGluGlyG......GlnGlnHisHisLeuGlyGly.AlaArgGln  239
1677 GCA...........GAGGCCCGCAGAGCGCGGGTGGGAGGTGGGAGCAGGCC  1637
      |||              :::|||  :::  |||  |||  ::
 240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl  256
1636 ACTGCCTCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTAG  1587
      ||||||  :::||                    |||||||||||
 256 yLeuProGlySerPro..................GlyAlaProGlyThr.  266
1586 GGCCTGGCCAGGAAGCTGGTCATCAGGGTGTCCTCACTGCTAGCACCTCCA  1537
      |||||||  |||  :::  |||  |||
 267 ..ProGlyProGlnGlyLeu.ProGlySerPro................  276
1536 GTGTCCCCTCGGTATTTGGCAGGAACACCTGCTTCTCCGGTGGTAGAG  1487
      ||::||||::::|||           ||||| :::|
 277 ...............GlyAlaProGlyThrProGlyProGlnG  286
1486 GGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGTGAGAAGTGAACCCGG  1437
      ||  |||        |||   ||| ||||| |||||
 286 lyLeuPro...........GlySerProGlyAla.ProGlyThrProGl  298
1436 TGAGGGCGGCTGAAGCTGTCAACACGGCCACACTGTGGGACAGGCATGTG  1387
      :::  ||||||             :::  |||||||||||||
 298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyA  313
1386 GCACCGGCAG.............CCACAGGGAAAGCTG  1362
      |||:::|||              ||||||| |||
 313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu  328
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCGAATCGCTGACCAGCCGG  1312
      |||  ||:||||||||:::              |||  |||
 329 ProGlySerPro.....................GlyAlaProGlyThr  337
```

```
;   REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 781-1989
;     TELEFAX:  (415) 398-3249
;     TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  31:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  762 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-397-633A-31

alignment_scores:
        Quality:   395.00         Length:       854
        Ratio:      1.082         Gaps:          59
Percent Similarity:  42.740     Percent Identity:  30.211

alignment_block:
US-09-030-606-110/rev x US-08-397-633A-31

Align seg 1/1  to: US-08-397-633A-31  from:  1  to: 762

2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGG 2298
     |||::|:|||||         :::|||::  ::|:|||::  ||
  76 GlySerProGly..........AlaProGlyThrProGlyGluGl 87

2297 TGTGTCT...CAACTAGGAGGCTAGCTGTTAACCCTGACCTGGGTAATC 2251
     |||:||||||        :::  :::|:||||:   |||:::|:||
  87 yGlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerP 104

2250 CACCTGCA.......GAGTCCCCGCATTCCAGTGCATGGAGCCCTTCT 2210
     ||  ||        ::||:|||         |||::|||||    |||||
 104 roGlyAlaProGlyThrProGlnGlyProGlyLeuProGlySerPro... 119

2209 GGCCTCCCTGTATAAGTCCAGACTGAAACCCCCTTGGAAGGCCTCCAGTC 2160
     ||| |||        |||||||||       |||   ::||||||
 120 GlyAlaPro.........GlyThrProGlyPro...GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGAGAGGAGAGGAGGACGCCCAGCCCCAG 2110
     ||||||||||||        :::   |||||||||||
 131 oGlySerPro..........GlyAlaProGlyThrProGlyProGlnG 144

2109 CTGTGCAGTGTACGGACCTCAGCACCCACCACAGGGT......GGCAGCAGAGAG 2066
     ::|::     |||   :::||  :::|||      ||||||
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159

2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCGGCAGCCCCA 2016
     |||::::::|||
 160 GlnHisHisLeuGly..... 164

2015 TGGGGCTAACAGGAGCGGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCAC 1966
     ||||||:::::|||:||||||:||   |||  |||:::||
 165 .GlyAlaArgGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180

1965 CCCAATGTGCTGGAAGTTTTCTACCGTGAGTATTTGGCCAAGTCGCTCTT 1916
     ||||                              :::
 180 rPro...........................................GlyP 183

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGGCGACCAGGCCAGGCCTG 1866
     ||| |||||                          ::|||||
 183 roGlnGlyLeuPro...........................GlySerProGlyAla 192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGACGTGGACAATGGAG 1816
     |||||        |||
 193 ProGlyThrProGlyGln.............................. 199

1815 CCCATAAACAGGATGGGGCCACTGGGACAGCAGGAAGGCACTATCCAG 1766
     ||||::::::  |||    :::
 200 ............GlyLeuProGlySerProGlyAla.......ProG 209
```

```
546 AAGGGCCGGCGGCGGCCATAGCGTCCACGCCAGTGGTCACTGCTGCTGAGCC 497
    |||
611 yLeuProGly...................................SerP 616

496 TAGGAGCGGGACACAGACCAGGCCCAGCACTGGACCAATGCCCAGCACCA 447
    |||||||||
616 roGlyAlaProGlyThrProGlyPro...................... 624

446 TGGTCATGAACTTCTCCTCTACCCCCACTTCCAGCAGCAGCAGAGGCGGCACA 397
    :::  |||||    |||:::|||
625 ...........GlnGlyLeuProGlySerProGlyAla........... 633

396 TAGGTGATGCCTGCGGGCCAAAACACACCTCCAGGCCAAAGGTTAGCCAGGTT 347
    |||::|||||   ||| |||   |||
634 ..........ProGlyThrProGlnGlyProGlnLeuProGlyLys 645

346 GACCAGCAAGAGCTGGGCTTCCGGT........................ 321
    |||::::  |||    |||
645 erProGlyAlaProGlyThrProGlyGluGlyGlnHisHisLeuHisGly 661

320 ........GCCGCAGCAGGCGGCTCACCCACAGCCTCTGGACCATA 283
    |||::::
662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678

282 GTGGGCCAGGCGGG...TAGGGCTCAGGGGCCGTTCAGGCACTCCAGAA 236
    ||||||  |||    ::|||  |||   |||
678 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 694

235 CTGCTTCGTCTCCGCCTCTGCCTCCAGAAGTCGCCGCCTCTCCTCCTGCTG 186
    |||  ::::|||  |||      ::::|||
695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly...... 709

185 CCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTCTGCCAGCCCTTTG 136
    |||||||||       :::|||:::|||||
710 .........ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722

135 GTGCCGGTCCAGCTTCTCAGCCATGCTCAACACCTGCTGCTGTGGGGCA 86
    ||   :::::|||   |||:::      |||
722 hrProGlyGluGlyGlnHisHisLeuGly...........GlyAla 734

 85 CCTCAGTGGGGACACGTCTCATCA.........CTCAGATCCTGGCCGA 46
    |||  |||:::|||       :::  |||||||||||
735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-31

seq_documentation_block:
; Sequence 31, Application US/08397633A
; Patent No. 5773577
; GENERAL INFORMATION:
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
;   TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
;   NUMBER OF SEQUENCES: 105
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/397,633A
;   FILING DATE:
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rowland, Bertram I
;   REGISTRATION NUMBER: 20,015

```
1765 GATGCGGAGGTCCAGGCAGATGCCCCGGC...........CCGGAACCA 1728
     |||   |||||||                          ||| :::
 209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225

1727 CCCTGGCCTCCGTGGGCTCACCCACCACCACACGTACGAGACATCACAG 1678
     ||||||| |||                    ::||||||       :::
 225 rProGlyGludly.....GlnGlnHisHisLeuGlyGly.AlaArgGln 239

1677 GCA.......GAGGCCCCGCAGAGCGCGGGTGGAGGTGGAGGAGGCC 1637
             :::||   ||    :::   |||   |||   |||
 240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 256

1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTAG 1587
     ||||||||    ::||||                ||||||||||
 256 yLeuProGlySerPro.............       GlyAlaProGlyThr. 266

1586 GGCCTGGCCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCA 1537
     |||||    |||                               |||
 267 ..ProGlyProGlnGlyLeu.ProGlySerPro............ 276

1536 GTGTCCCCTCGGTATTTGGGCAGGAACCACCTGCTTCTCCCGGTGGTAGAG 1487
     |||:::|||||:::||||||:::||||||        :::|
 277 ...........GlyAlaProGlyThrProGlyProGlnG 286

1486 GGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGTGAACCCGG 1437
     ||  |||        |||||  ||   ||||   ||||
 286 lyLeuPro........GlySerProGlyAla.ProGlyThrProGl 298

1436 TGAGGGCGGCTGAAGCTGCCACACGTGCCACACTGTGGACAGGCATGTG 1387
     ||||        :::     |||||       ::  ||||||
 298 yGluGly......GlnGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyA 313

1386 GCACCGGCAG.....................CCACAGGGAAAGCTG 1362
     ::|||                            |||||||| |||
 313 spValGlySerProGlyAlaProGlyThrProGlnGly...Leu 328

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGG 1312
     |||                     ::|||||||||   |||
 329 ProGlySerPro...........       GlyAlaProGlyThrGlyTh 337

1311 TCCATGACCAGGAGAGACCAGGGAGAGTGGCCACTGCAGGAACA..... 1266
     ||||||||||       |||:::       ||||||
 337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354

1265 .......................GCCCCAGGCTGCCCATCCGAA 1245
                            |||||||||   |||:::|||
 354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTTGTCTCCGGGCCTCGGTGCCCGGCTCAGCTCTGGGC 1195
     ::|||:::::||   |||:::::::||
 371 GlyGlnGlnHisHisLeuGlyLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe 387

1194 A......CGCCCTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG 1151
     :      ||||||||   ||| |||:::::
 387 rProGlyAlaProGlyThr..ProGlyProGlnGly....... 398

1150 AGGGCGGCCCAGGTTCC...GGAAAGCCAAGCCGGGCCCGGC......... 1101
     |||  ||||||||  ||   |||     |||||
 399 .........LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGGCGGCAGGGTGCGGGGCCATGGGCAGCGGCACAGCTGGTGCAGCCGGAAGC 1051
     |||||||||||||||||                |||
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGCGGCCCAGGTTC...GGAAGCCAAGCCGGGCCCGGC.......... 1014
     |||  |||||||||   ||||||||||||||
 427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444

1013 ..........................ATGGACAGCAGTGGG 999
                               ::  ||||
 444 lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerProGly 460

 998 GCCGACA...AGGAGGGGGGCCGACAGCCCCTTCTGCTGCTCGGTGGGCCCC 952
     ||||||
```

```
 461 AlaProGlyThrProGlnGlyLeu.......ProGlySerPr 474

 951 AGCGGCTGCCTCCTCAGCCACCACCAGCAGTGTGGCTGCTGCTACGGCAGGTGAGGAA 902
     |::: |||                     |||
 474 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 491

 901 GATGAGGGTGAGCAGGCCAAAGAGGCCACTCCT.............CCT 867
     |||                     |||
 491 laProGlyThrProGlnGlyLeuProGlySerProGlyAlaPro 507

 866 GGGTGCCCAGTAGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGC 817
     |||  |||||::: |||           ::: |||||
 508 GlyThrProGlyGluGlyGlnHisHisLeuGlyGlyAlaArgGlnAl 524

 816 AGGA.......GGTAGCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATA 773
     :::                              :::||
 524 aGlyAspValGlySerProGlyAlaPro........... 533

 772 GACAGAGTAGGCCTGGCCGACAGTGGT...CCGGGT...CCCGGAAGAGGT 729
     ::|||   |||:::||   |||||   ||||||
 534 ....GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 548

 728 CAGAGAGCCAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCC 679
     |||||||||                    :::|||||
 549 GlyThrProGlyProGlnGlyLeuProGlySerProGly......AlaPr 563

 678 AGCAGCCCCACGCCCAGGATGAGCAGTGCCAGTCCAGGGGCCTGGGATC 629
     :: :::|||   |||           ::|||||||||
 563 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.......P 578

 628 CGGGCACAGCAGCCCTGCTAGCCAGCCGGCCCTTGGATGAGAAAGA... 582
     |||||||||   :::          :::|||   |||   |||
 578 roGlyThrProGlyGluGlyGlnHisHisLeuGlyGlyAlaArgGln 594

 581 .............GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATG 547
     ||||||||||                   |||
 595 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 611

 546 AAGGGGCCGGCGGCGGCCATAGCGTCCACGCCCAGTGTCACTGGCTGGAGCC 497
     ||||
 611 yLeuProGly.................................SerP 616

 496 TAGGAGCGGGACACAGACACCAGGCCCAGCACTGGACCAATGCCCAGCACCA 447
     |||||   |||||||||
 616 roGlyAlaProGlyThrProGlyPro.................. 624

 446 TGGTCATGAACTTCTCCTCTACCCCCCACTTCCAGCAGCAGAGGCGGCACA 397
     ||||       |||:::|||
 446 ...............GlnGlyLeuProGlySerProGlyAla....... 633

 625 ..........GCCCCGCAGCAGCGGCTCACCCAGCCTCTGGACCCATA 347
     :::      |||||||||       ||||||
 396 TAGGTGATGCCTGCGGGCCAAACACACCTCCAGGCCCAAAGGTTAGCAGGTT 347

 634 ...................ProGlyThrProGlyProGlnGlyLeuProGlyLeuProGlyS 645
     |||:::|||||
 662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678

 282 GTGGGCCAGGCGGG...TAGGGCTCAGGGGCCGTTCAGGCACTCCAGAA 236
     ||||||| |||       |||||
 346 GACCAGCAAGAGCTGGGCTTCCGGT................... 321

 678 roGlyProGlnGlyLeuProGlyGluGlyGlnHisHisLeuGly 694
     ||||||       |||||
 645 erProGlyAlaProGlyThrProGlyGluGlyGlnHisHisLeuGly 661

 235 CTGCTTCGTCTCCGCTCGTCTCCAGAAGCTGCGGCCCTCCTCCTTGCTG 186
     |||:::|||||                   :::|||
 320 ............GCCCCGCAGCAGCGGCTCACCCAGCCTCTGGACCCATA 283

 695 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly...... 709
     |||   |||
 662 GlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.P 678

 185 CCGGCCAACTGCCTAGGAATCAGCCAGGCCGCCCATTTCTGCCAGCCCTTTG 136
     |||:::::     |||||||||::::
```

```
alignment_scores:
                 Quality:  393.00          Length:   852
                   Ratio:    1.059            Gaps:    59
    Percent Similarity:   43.545    Percent Identity:  30.516

alignment_block:
US-09-030-606-110/rev x US-08-707-237A-84   .

Align seg 1/1   to: US-08-707-237A-84   from: 1  to: 761

2347 GGGAAACCAGGTGACTGAGTTTATTCAGTCCCAAAAACCCTCTCTAGG 2298
     |||::|||||||        |||:|||        |:::||:::  ||
  75 GlySerProGly..............AlaProGlyThrProGlyGluGl 86

2297 TGTGTCT..CAACTAGGAGGCTAGCTGTTAACCTGAGCCTGGGTAATC 2251
     |  :::   ::|||||||||      |:: ::::|||:||:|  ||
  86 yGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 103

2250 CACCTGCA..........GAGTCCCGCATTCCAGTGCATGGAGCCCTTCT 2210
     || ||     :|||::|||||:|||        :|:
 103 roGlyAlaProGlyThrProGlyProGlyLeuProGlySerPro... 118

2209 GGCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCTCCCAGTC 2160
     ||| |||         |||||||||        |||  :::|||:||
 119 GlyAlaPro.......GlyThrProGlyPro...GlnGlyLeu...Pr 130

2159 AGGCAGCCCTAGAGACTGGGAGGAGAGGACGCCCAGCCCCCAG 2110
     ||||||||         |||||||||  |||      .
 130 oGlySerPro..........GlyAlaProGlyThrProGlyProGlnG 143

2109 CTGTGCAGCTACGCACCTCAGCAGCACAGGGT......GGCAGCAGAGAG 2066
     :::  ::|||   ||    ::::|||     :::
 143 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 158

2065 CCACATTACTTTGCAGCAACAGAAACTGGCGCCCAGCCCGGCAGCCCCA 2016
     ........................................
 159 GlnHisHisLeuGly.......................... 163

2015 TGGGGCTAACAGGAGGCGGGAGCTGGGACCCAGTGAGCAGGCCCTCCAC 1966
     ||||::::|||||||:|||:|||:|||        |||  |||:::||
 164 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 179

1965 CCCAATGTGCTGGAAGTTTTCTACGCTGAGTATTTGGCCAAGTCGCCTCTT 1916
     |||                                ::|
 179 rPro...........................GlyP 182

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGGCCTG 1866
     ||| ||||||        |||||||::::        ::||||||
 182 roGlnGlyLeuPro.............GlySerProGlyAla 191

1865 CGGCAGACACCATATAGGCAGCAGTGACAGACTGGCTGAGCTGGACAATGAG 1816
     |||                                  |||
 192 ProGlyThrProGlyProGln............. 198

1815 CCCATAAAACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCACTATCCAG 1766
     |||||              |||  ||||||||||        |||
 199 .............GlyLeuProGlySerProGlyAla......ProG 208

1765 GATGGGGAGGTCAGGCAGAGATGCCCCGGC............CCGGAACCA 1728
     ||  |||||||                            |||
 208 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 224

1727 CCCTGGCTCCGGTGGGCTCACCCACCACCACGTACGGAGACATCACAG 1678
     ||||| |||                ::||||||        |||:::|||
 224 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlaLysGln 238

1677 GCA........GAGGCCCCGCAGAGCCGCGGGTGGAGGTGGAGGACCAGGCC 1637
     |||        :::|||    |||        ||| |||     :::
 239 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 255

1636 ACTGCCTCCAGCACCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTAG 1587
```

```
710 .........ProGlnGlyLeuProGlySerProGlyAla...ProGlyT 722

135 GTGCCGGTCCAGCTTCTCAGCCATGCTCAACACCTGCTGCTGTGGGGCA 86
    |||  :::::|||        |||:::        ||||||
722 hrProGlyGluGlyGlnGlnHisHisLeuGly...........GlyAla 734

 85 CCTCAGTGGGACACGTCTCATCA.......CTCAGATCCTGGCCGA 46
    |||    |||  |||:::||        ::: |||||||||||
735 ArgGlnAlaGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-84  .

seq_documentation_block:
; Sequence 84, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
;   APPLICANT:  Ferrari, Franco A.
;   APPLICANT:  Capello, Joseph
;   APPLICANT:  Crissman, John W.
;   APPLICANT:  Dorman, Mary A.
;   TITLE OF INVENTION:  METHODS FOR PREPARING SYNTHETIC
;   TITLE OF INVENTION:  REPETITIVE DNA
;   NUMBER OF SEQUENCES:  108
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;     STREET:  Four Embarcadero Center, Suite 3400
;     CITY:  San Francisco
;     STATE:  California
;     COUNTRY:  United States
;     ZIP:  94111-4187
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/707,237A
;     FILING DATE:  03-SEP-1996
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/175,155
;     FILING DATE:  29-DEC-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/053,049
;     FILING DATE:  22-APR-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/609,716
;     FILING DATE:  06-NOV-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/269,429
;     FILING DATE:  09-NOV-1988
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 07/114,618
;     FILING DATE:  29-OCT-1987
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 06/927,258
;     FILING DATE:  04-NOV-1986
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Trecartin, Richard F.
;     REGISTRATION NUMBER:  31,801
;     REFERENCE/DOCKET NUMBER:  A-55186-10/WHD
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 781-1989
;     TELEFAX:  (415) 398-3249
;     TELEX:  910 277299
; INFORMATION FOR SEQ ID NO:  84:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  761 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  unknown
;     TOPOLOGY:  unknown
;   MOLECULE TYPE:  protein
US-08-707-237A-84
```

us-09-030-606-110.rai

```
255 yLeuProGlySerPro..................GlyAlaProGlyThr. 265
    ||||||              ::||||                ||||||||||
1586 GGCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCA 1537

266 ..ProGlyProGlnGlyLeu.ProGlySerPro.............. 275
    ||||||    :::   |||   ||||||
1536 GTGTCCCCTCGGTATTTGGGCAGGAACACCTGCTTCTCCCGGTGGTAGAG 1487

276 .................GlyAlaProGlyThrProGlyProGlnG 285
    ||||::||||||:||||
1486 GGAGGCCAGTGTGTAGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGG 1437

285 lyLeuPro..........GlySerProGlyAla.ProGlyThrProGl 297
    || |||          |||||   :::  ||  |||||
1436 TGAGGCGGCTGAAGCTGTCACCACGGCCACACTGTGGGACAGGCATGTG 1387

297 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 312
    ||||::|          :::   |||||
1386 GCACCGGCAG..................CCACAGGGAAAGCTG 1362

312 spValGlySerProGlyAlaProGlyThrProGlnGlyLeu...Leu 327
    :::|||        |||||||            |||
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCCAGCCGG 1312

328 ProGlySerPro.................GlyAlaProGlyTh 336
    |||  ||||||                ::|||||:::
1311 TCCATGACCAGAGAGAAGACCAGGAGATGGCGACTGCAGGAACA.... 1266

336 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 353
    |||  :::       ||||||
1265 ..................GCCCCAGGCTGCGCCCATCCGAA 1245

353 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 369
    ||||||||     |||   |||::||
1244 CGCTTCATCATAGTGTCTCCCGGGCTCCGGTCCCCGGCTCAGCTCTGGGC 1195

370 GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 386
    ::||||           :::   ||::|:   |||:|
1194 A......CGCCCTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG 1151

386 rProGlyAlaProGlyThr...ProGlyProGlnGly........ 397
    :   ||||||||   ||:::   ||||
1150 CGTGAAGGTCATCAGTGCCATCCAGCTGCACAGTCCAGACGCCACGAAGAGCC 1101

398 ...........LeuProGlySerProGlyAlaProGlyThrProGly 409
    ::||||:|||      |||       |||       ||||||
398

1100 GGGCCAGGGTGCGGGGCATGCGGCAGCAGCTGGTGCAGCGCGGGGAAGC 1051

410 AlaProGlyThrProGlyLeuProGlyLeuProGlySerProGlyProGl 426
    ::   |||:::||      ::::|||     |||
963 TCGGTGGGG...................CCCAGCGCTGC 944

443 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 459
    :::|||      ||||:::||||

1050 AGGGGCGCCCAGGTCC..GGAAAGCCAAGGCGGCGCCCGGCAT.GGACAGC 1005

426 nGlyProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 443
    ||| :::||||||   |||:::  |||||||

1004 AG........TGGGGCGGACAAGGAGGGGGCCGCCGCACAGCCCTTCTGCTGGC 964

459

1004 

476 AlaProGlyThrProGlyProGlnGlyLeuProGlyLeuProGlySerProGlyProGlyAla.P 476

943 CTCCTCAGCCACCACCAGTGTGGCTGCTACGCAGGTGAGGAGGAAGATGAGGG 894

476 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 492

893 TGAGCAGGCCAAAGAGCACTCCT.............CCTGGGTGCCC 859

493 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 509

858 AGGTAGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA.... 813
```

```
509 oGlyGluGlyGlnGlnHisHisLeuGlyGlyGlyAlaAlaLysGlnAlaGlyAspV 526
    ||||||||||||||||||||::||
812 ..GGTAGCCCAGGCAGCCCCCAAGACTGATCATGAGGCATAGACAGAGT 765

526 alGlySerProGlyAlaPro.......................:::Gly 533

764 AGGCCTGGCGACAGTGGT...CCGGGT...CCCGGAAGAGAGTCAGAGAGC 721
534 ThrProGlyProGlnGlyLeuProGlyLeuProGlySerProGlyAlaProGlyThrPr 550

720 AGGGGCCCCAGTGGAGTGAAGCACACCTGCCACAGAAGTCCAGCAGCCC 671
550 oGlyProGlnGlyLeuProGlySerProGly.....AlaProGlyThrP 565

670 CACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCTCGGATCCGGCACA 621
565 roGlyProGlnGlyLeuProGlySerProGlyAla......ProGlyThr 579

620 GCAGCCCTGCTAGCCAGCCGCCCCTTGGGATGAGAAAGA.......... 582
580 ProGlyGluGlyGlnGlnHisHisLeuGlyGlyGlnAlaGlyGlnAlaGlyAs 596

581 ....GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATGAAGGGCCG 539
596 pValGlySerProGlyAlaProGlyThrProGlnGlyLeuProG 613

538 GCGGCGGCCATAGCGTCCACCGCCAGTGTCACTGCCTGAGCCTAGGAGCG 489
613 ......................................SerProGlyAla 617

613 ly..
488 GGACACAGACAGGCCCAGCCACCACTGGACCAATGCCCAGCCAGCACCATGGTCATG 439

618 ProGlyThrProGlyPro....................... 623

438 AACTTCTCCTCTACCCCACTTCCAGCAGCAGCAGAGGCGGCACATAGGTGAT 389

624 ....GlnGlyLeuProGlySerProGlyAla........ 632

388 GCCTGCGGCAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339

633 ........ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 646

338 AGAGCTGGGCTTTCCGGT 321

647 AlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLy 663
321 ly...............

320 ....GCCGCAGCAGGCGGCTCACCCACAGCCTCTGGACCATAGTGGGCCA 275

663 sGlnAlaGlyAspValGlySerProGlyAla...ProGlyPro 679

274 GGCTCAGGG...TAGGGCTGGGGCCGTTCAGGCACTCCAGAACTGCTTCG 228

680 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl 696

227 TCTCGGCTCTGCTCCAGAAGCTGCGCGCCCTCCTCCTTGCTGCCGCCAAC 178

696 yLeuProGlySerProGlyAlaProGlyThrProGly............. 708

177 TGCCTAGGAATCAGCCAGCGGCGCCCATTCTGCCAGCCCTTGGTGCCGGT 128

709 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 723

127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTG 78
724 GluGlyGlnGlnHisHisLeuGlyGly.............GlyAlaLysGlnAla 736

77 GGGACAGCTCATCA.........CTCAGATCCTGGCCGA 46
736 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 749
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-114

seq_documentation_block:

```
; Sequence 114, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
;   APPLICANT: CAPPELLO, Joseph
;   APPLICANT: FERRARI, Franco A.
;   TITLE OF INVENTION: High Molecular Weight Collagen-Like
;   TITLE OF INVENTION: Protein Polymers
;   NUMBER OF SEQUENCES: 135
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;     STREET: 4 Embarcadero Center, Suite 3400
;     CITY: San Francisco
;     STATE: California
;     COUNTRY: USA
;     ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/642,255
;     FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: ROWLAND, Bertram I.
;     REGISTRATION NUMBER: 20,015
;     REFERENCE/DOCKET NUMBER: A55556-3/BIR
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 494-8700
;     TELEFAX: (415) 494-8771
;     TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 114:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 762 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-642-255-114
```

```
alignment_scores:
       Quality:  393.00          Length:   852
         Ratio:    1.059            Gaps:    59
Percent Similarity:  43.545    Percent Identity:  30.516

alignment_block:
US-09-030-606-110/rev x US-08-642-255-114

Align seg 1/1  to: US-08-642-255-114  from: 1  to: 762
```

```
2347 GGGAAACCAGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTTAGG 2298
     |||::|||||                ::|||:::|| ::|||:::|
  76 GlySerProGly..........AlaProGlyThrProGlyGluGl 87

2297 TGTGTCT...CAACTAGGAGGCTAGCTTAAACCCTGAGCCTGGGTAAATC 2251
     || :::: ::|||||||||||      ::: :||||||::::|
  87 yGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerP 104

2250 CACCTGCA...........GAGTCCCGCATTCCAGTGCATGGAGCCCTTCT 2210
     || |||||                  ::|||:|||||||:
 104 roGlyAlaProGlyThrProGlnGlyLeuProGlySerPro... 119

2209 GGGCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTC 2160
     |||  |||         ||| ::|||||
 120 GlyAlaPro.......GlyThrProGlyPro...GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGGAGGAGGAGGAGCCGCCCCCAGCCCCCAG 2110
     |||||||||            ||||||||| |||
 131 oGlySerPro.........GlyThrProGlyAlaProGlyThrProGlnG 144

2109 CTGTGCAGCTACGCACCTCAGCACGCACAGGGT.....GGCAGCAGAGAG 2066
```

```
        ::::|| ..   :::||   ::||| ...     :::
144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159
     ||||||||||||            .............
2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCCGGCAGCCCCA 2016
     |||||||||||| ||||||
160 GlnHisHisLeuGly.......................GlySerProGlyAla 164
     |||||::::||:|| ||:|||:|
2015 TGGGCTAACAGAGAGCGGGAGCTGGACCCAGTGAGCGAGGCCTCCAC 1966
     |||||:::::||::|||||:||| |||  |||:::||
165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyThr 180
     ||||
1965 CCCAATGTGCTGGAAGTTTTCACGCTGGAGTATTTGGCCAAGTCGCTCTT 1916
                                     :::
180 rPro.........................................GlyP 183
     ||||
1915 GTCAAATACTACCTGTGTAGCAAGTAAATGGCGACCACAGACCCAGGCCTG 1866
       |||  |||||       ::|||||||
183 roGlnGlyLeuPro.....................GlySerProGlyAla 192
     :::||||||
1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGAG 1816
     ||||||||
193 ProGlyThrProGlyGln................................ 199
     |||||
1815 CCCATAAACAGGATGGGGCCACCTGGGACAGCAGGAAGGCACTATCCAG 1766
                      |||  |||||||:::   |||
200 .....................GlyLeuProGlySerProGlyAla......ProG 209
                                          ||||
1765 GATGGCGAGGTCCAGGCAGATGCCCCGC...............CCGGAACCA 1728
     ||      ::|||||      ||| :::
209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225
     |||:::|||
1727 CCCTGGCCTCGGTGGCTCACCCACCACCACGTACGGAGACATCACAG 1678
     |||||||        ::|||||  |||  :::||||
225 rProGlyGluGly.....GlnGlnHisHisLeuGlyGly.AlaLysGln 239
     |||
1677 GCA.......GAGGCCCCGCAGAGCGCGGGTGGAGGTGGAGGCAGGCC 1637
     |||        ::|||  |||  :::
240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyGlnGl 256
                      |||||||||
1636 ACTGCCTCCAGCACCCACGTGTCCATTAGGGAGGGAGCTCCAGGCTTAG 1587
     ||||| ::|||
256 yLeuProGlySerPro................GlyAlaProGlyThr. 266
     |||||||
1586 GGCCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCCACTGCTAGCACCTCCA 1537
     |||||  ::: ||||| |||
267 ..ProGlyProGlnGlyLeu.ProGlySerPro................. 276
     |||||:||:|||:::|||||  ::|||
1536 GTGTCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCGGTGGTAGAG 1487
     ||||:||:|||::::|||:: :::|
277 ...........................GlyAlaProGlyThrProGlyProGlnG 286
     ||||| |||      |||||
1486 GGAGGCCAGTGTGTGAGGCGAGATCTGCAGGGTGAGAAGGTGAACCCGG 1437
     || |||              ::|||||
286 lyLeuPro.........GlySerProGlyAla.ProGlyThrProGl 298
     ||||| |:|
1436 TGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGTGGACAGGCATGTG 1387
     |||||| ::: ||||||
298 yGluGly.....GlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyA 313
     ||||||
1386 GCACCGGCAG.................CCACAGGGAAAGCTG 1362
     ::|||      ||||||||||     |||
313 spValGlySerProGlyAlaProGlyThrProGlnGly...Leu 328
     |||
1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCCACCAGCCGG 1312
     ||| |||                ::|||||
329 ProGlySerPro....................GlyAlaProGlyThr 337
     |||
1311 TCCATGACCAGGAGAGACCAGGAGATGGCCACTGCAGGAACA...... 1266
     ||||| :::                 :: ||||
337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354
     |||||||            ||||
1265 ................................GCCCCAGGCTGCCCATCCGAA 1245
     |||::||||
```

Wed Sep 29 14:26:53 1999

354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTGTCTCCGGGCCCTTCGGTCCCGGCTCAGCTCTGGGC 1195
     :::::||| :::::||| ::   :::::|||
371 GlyGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlyLyse 387

1194 A......CGCCTTGGTACAGCCCCTCGCCACGAAATCCGTGTAAAACAG 1151
        ||||||||| || |||:::::|
387 rProGlyAlaProGlyThr...ProGlyProGlnGly............ 398

1150 CGTGAAGGTCATGAGTGGCATCCAGCTCGCACAGCTCAGCCACGAAGAGCC 1101
     |||||||:||| |||
399 ....LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGGCGCAGGGTGCGGGGCATGCGCGACACAGCTGGTCAGCCCGGGAAGC 1051
     ||| ::
411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGGCGCCCAGGTTCC...GGAAAGCCAAGCGGGGCCCGGCAT.GGACAGC 1005
     ||| ||||||||| ||| |||::::: |||||||
427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444

1004 AG........TGGGGGACAAGGAGGGGGCCGACAGCCCTTGCTGGC 964
     || ||| |||:::::||||
444 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 460

963 TCGGTGGGG................................CCCAGCCGCTGC 944
    ::: ||
461 AlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.P 477

943 CTCCTCACCACCAGCAGTGTGCTGCTGTACGCCAGGTGAGGAAGATGAGGG 894
    ||| ||| |||
477 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 493

893 TGAGCAGGCCAAAGAGCCACTCCT................CCTGGGTGCCC 859
    ||| :::|||| ||
494 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510

858 AGGTAGGGGGGCCACTGGGTCTCCCAGTCAATGCAGCCAGGCAGGA.... 813
    ||||:||||||| ||||||||
510 oGlyGluGlyGlnHisHisLeuGlyGlyAlaLysGlnAlaLysGlyAspV 527

812 ..GGTAGCCCAGGCAGCCCCAAGACTGATCATGAAGGCATAGACAGAGT 765
    ||||||||||:::|| ||||||||
527 alGlySerProGlyAlaPro...........................Gly 534

764 AGGGCCAGGGACCAGTGGT...CCGGGT...CCCCGAAGAGGTCAGAGAGC 721
    ||||| ||| |||||| ||||||||
535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551

720 AGGGGCCTTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCC 671
    |||||| :::::|||||
551 oGlyProGlnGlyLeuProGlySerProGly......AlaProGlyThrP 566

670 CACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATCCGGCACA 621
    |||||||:||||||| ::|||||||
566 roGlyProGlnGlyLeuProGlySerProGlyAla......ProGlyThr 580

620 GCAGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGA.......... 582
    :::||||:::||||| ::
581 ProGlyGluGlyGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAs 597

581 ....GGCTCAGCAGGATGCCCAAGACAGTG...CCCAGATGAAGGGCCG 539
    |||||| |||:::||| ||| |||
597 pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 614

538 GCGCGGCGGCCATAGCGTCCACGCCAGTGCTCACTGGCTGAGCCTAGGAGCG 489
    ||| ||| |||||
614 ly............................SerProGlyAla 618

488 GGACACAGACCAGGCCCACTGACCAATGCCCAGCCACCATGGTCATG 439
    |||||||||
619 ProGlyThrProGlyPro............... 624

438 AACTTCTCCTCTTACCCCCACTTCCAGCAGCAGAGAGGCGGCACATAGGTGAT 389
    ::: |||||| |||:::||| :::||
625 ....GlnGlyLeuProGlySerProGlyAla.................. 633

388 GCCTGCGGCCAAACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339
    |||:::|||||| ||| ||| |||:::|||
634 ....ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647

338 AGAGCTCGGCGTTTCCGGT.............................. 321
    |||
648 AlaProGlyThrProGlyGluGlyGlnHisHisLeuGlyGlyAlaLy 664

320 ....GCCGCAGCAGGCGGCTCACCCACCGCGTCTGGACCATAGTGGGCCA 275
    ||||| :::||||||
664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680

274 GGCGGG...TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAACTGCTTCG 228
    ||| ||||||||| ||| ||| |||||||
681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl 697

227 TCTCGGCTCTGCTCACGAAGCTGCGGGCCTTCTCCTTGCTGCCGCCAAAC 178
    |||||||| ||| :::|||
697 yLeuProGlySerProGlyAlaProGlyThrProGly.............. 709

177 TGCCTAGGAATCAGCCAGGCGCGCCCATTCTGCCAGCCCTTTGGTGCCGGT 128
    |||:::||| |||||||| ::: |||
710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724

127 CCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCCACCTCAGTG 78
    :::::||| |||||||
725 GluGlyGlnHisHisLeuGlyGly..............GlyAlaLysGlnAl 737

77 GGGACACGTCTCATCA........CTCAGATCCTGGCCGA 46
    ||| |||:::||| |||||||||
737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-26

Wed Sep 29 14:26:53 1999

us-09-030-606-110.rai

```
;     LENGTH:  762 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
US-08-397-633A-26

alignment_scores:
          Quality:    393.00           Length:       852
          Ratio:       1.059            Gaps:          59
Percent Similarity:  43.545     Percent Identity:  30.516

alignment_block:
US-09-030-606-110/rev x US-08-397-633A-26

Align seg 1/1    to: US-08-397-633A-26    from:  1   to:  762
```

```
2347 GGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTAGG 2298
      |||::||||||                   ||||||  ::::||:::: ||
  76 GlySerProGly..................AlaProGlyThrProGlyGluGl  87

2297 TGTGTCT..CAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATC 2251
      |   ||  ::|||||||||      ::  :|||||||:::|||::::|
  87 yGlnGlnHisLeuGlyGlyAlaLysGlyAlaLysGlnAlaGlyAspValGlySerP 104

2250 CACCTGCA.......GAGTCCCGCCATTCCAGTGCATGGAGCCCTTCT 2210
      ::|||         ||      |||||||                   |||
 104 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPro... 119

2209 GGCTCCCGTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTC 2160
      |||     |||  ||||      ::|||||
 120 GlyAlaPro........GlyThrProGly...GlnGlyLeu...Pr 131

2159 AGGCAGCCCTAGAGACTGGGGAGAGAGGACGCCCCAGCCCCCAG 2110
      ||||||||        ||||||||  |||
 131 oGlySerPro.........GlyAlaProGlyThrProGlyProGlnG 144

2109 CTGTGCAGTGACGCACCTCAGCACAGGGT......GCAGCAGAGAG 2066
      ::|||    ::|||  |||:  ::||:::||
 144 lyLeuPro...GlySerProGlyAlaProGlyThrProGlyGluGlyGln 159

2065 CCACATTACTTTGGCAGCAACAGAGAAACTGCGGCGGCCAGCAGCCCCA 2016
      |||:||||||         ::|||||||
 160 GlnHisHisLeuGly.................GlySerProGlyAla 192

2015 TGGGGCTAACAGGAGCGGGGAGCTGGGACCCCAGTGAGGCAGGCCCTCCAC 1966
      ||||||:::||::|||   ||| |||::||
 165 .GlyAlaLysGlnAlaGlyAspValGlySerPro..GlyAlaProGlyTh 180

1965 CCCATAAACAGGGATGGGGCCACTGGGACCAGCAGGAAGGCCACTATCCAG 1916
      :::||   ||||:::||   |||::
 180 rPro..........................GlyP 183

1915 GTCAAATACTACCTGTGTAGCAAAGTAAATGGCCAGACCCAGGCCTG 1866
      ||   |||||                     ::|||||||
 183 roGlnGlyLeuPro.................GlySerProGlyAla 192

1865 CGGCAGACACCATATAGGCAGTGACAGACTGGCTGACAATGGAG 1816
      |||||  |||
 193 ProGlyThrProGlyGln.................. 199

1815 CCCATAAACAGGGATGGGGCCACTGGGACCAGCAGGAAGGCACTATCCAG 1766
      |||||:::::::||  |||:::||                 ||||
 200 ........GlyLeuProGlySerProGlyAla......ProG 209

1765 GATGGCGGAGGTGCCAGGCAGATGCCCCGGC..........CCGGAACCA 1728
      |||||  |||   ||  |||                      ||| :::
 209 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro.GlyTh 225

1727 CCCTGGCCTCGGTGGGCTCACCCACCACCACCACGCGGAGACATCACAG 1678
      |||||| |||                 ::|||||| ||| :::||:||
 225 rProGluGluGly.....GlnGlnHisLeuGlyGly.AlaLysGln 239
```

```
1677 GCA........GAGGCCCCCAGAGCGCGGGTGGAGGTGGAGCAGGCC 1637
      |||                   :::||  |||      |||  :::
 240 AlaGlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGl 256

1636 ACTGGCTCCAGCACCCACGTGCCATTAGGGAGGGAGCTCCAGGCTTAG 1587
      ||||||                            ||||||||
 256 yLeuProGlySerPro..............GlyAlaProGlyThr. 266

1586 GGCCTGGCGAAGAGCTGGTCATCAGGCTGTCCTCAGCTGCTAGCACTCCA 1537
      ||||| |||  :::  |||  |||  |||
 267 ..ProGlyProGlnGlyLeu.ProGlySerPro................ 276

1536 GTGTCCCCTCCGTATTGGGCAGGACACCTGCTTCTCCCGTGGTAGAG 1487
      |||::||||::::||||||  :::|
 277 ...................GlyAlaProGlyThrProGlyProGlnG 286

1486 GGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGTGAACCCGG 1437
      || |||   ||||||  |||||    |||::  ::|||||||
 286 lyLeuPro..........GlySerProGlyAla.ProGlyThrProGl 298

1436 TGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGGGACCAGGCCATGTG 1387
      ||||||                ::: :|||||      |||||
 298 yGluGly......GlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyA 313

1386 GCACCCGCAG........................CCACAGGGAAAGCTG 1362
      ::|||                               |||||| |||
 313 spValGlySerProGlyAlaProGlyThrProGlyProGlnGly...Leu 328

1361 CCACACTGGCCAAATAGACTGCTCAGTGCCGAATCGCCTGCCAGCCGG 1312
      |||     ::|||||||||||||                ::|||||||
 329 ProGlySerPro..................GlyAlaProGlyThr 337

1311 TCCATGACCAGAGAAGAGACCACCAGGAGATGGCCACTCAGGAACA...  1266
      |||   |||:::  |||||                   ::: |||
 337 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 354

1265 ................................GCCCCAGGCTGCCATCCGAA 1245
      ::|||                           |||||||::::|||
 354 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGlu 370

1244 CGCCTTCATCATAGTGTCTCCGGGCTCGGTGCCCGGCTCAGCTCTGGGC 1195
      |||:|                      :::|||||||||:::||:||
 371 GlyGlnGlnHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 387

1194 A.......CGCCTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG 1151
      :    |||||| |||   :::|||:::
 387 rProGlyAlaProGlyThr..ProGlyProGlnGly............. 398

1150 CGTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCAGCCGAAGAGCC 1101
      :::|||:::|||                 |||                :::
 399 ................LeuProGlySerProGlyAlaProGlyThrProGly 410

1100 GGCGAGGGTGCGGGGCATGGCAGCACAGCTGGTGCAGCCGGGGAAGC 1051
      |||
 411 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 427

1050 AGGGGCGCCCAGGTTCC...GGAAAGCCAAGCGGGGCCCGGCAT.GGACAGC 1005
      |||  ||||||||| |||  ::: |||||| |||||  :|||||
 427 nGlyLeuProGlySerProGlyAlaProGlyThrProGlyGluGlyGlnG 444

1004 AG.........TGGGGCGACAAGGAGGGGGCCGACAGCCCCTTCTGCTGGC 964
      :  ||                |||::|||   |||
 444 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySerProGly 460

963 TCGGTGGGG..............................CCCAGGGCTGC 944
     ::: |||                                 |||:::||
 461 AlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla.P 477

943 CTCCCAGCCAGTGGCTGCTCCACCGGTGCAGGTGAGGAAGAATGAGGG 894
     ||| |||                |||  |||::|||| |||
 477 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 493
```

Wed Sep 29 14:26:53 1999

737 aGlyAspValGlySerProGlyAlaMet.AspProGlyArg 750
    ||| |||||||||:::|| ||||||||||||||
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-126

seq_documentation_block:
; Sequence 126, Application US/08642255
; Patent No. 5773249
GENERAL INFORMATION:
    APPLICANT: CAPPELLO, Joseph
    APPLICANT: FERRARI, Franco A.
    TITLE OF INVENTION: High Molecular Weight Collagen-Like
    TITLE OF INVENTION: Protein Polymers
    NUMBER OF SEQUENCES: 135
    CORRESPONDENCE ADDRESS:
    ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
    STREET: 4 Embarcadero Center, Suite 3400
    CITY: San Francisco
    STATE: California
    COUNTRY: USA
    ZIP: 94111-4187
    COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/642,255
    FILING DATE:
    CLASSIFICATION: 435
    ATTORNEY/AGENT INFORMATION:
    NAME: ROWLAND, Bertram I.
    REGISTRATION NUMBER: 20,015
    REFERENCE/DOCKET NUMBER: A55556-3/BIR
    TELECOMMUNICATION INFORMATION:
    TELEPHONE: (415) 494-8700
    TELEFAX: (415) 494-8771
    TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 126:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 682 amino acids
    TYPE: amino acid
    STRANDEDNESS: double
    TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-642-255-126

alignment_scores:
        Quality: 370.50   Length: 831
        Ratio: 1.015   Gaps: 59
Percent Similarity: 43.923   Percent Identity: 29.964

alignment_block:
US-09-030-606-110/rev x US-08-642-255-126

Align seg 1/1   to: US-08-642-255-126   from: 1 to: 682

2314 AAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     ::|||||| |||||| ::||||:::::||
  12 GluAsnPro.....GlyValThrGlnLeuAsn...........ArgLe 22

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATTCCAGTGCATGGAGCC 2215
     |::| |||||||| :::||||| |:::||
  22 uAlaAlaHisProProPheAlaSerAspProMetGlyAlaProGlyThrP 39

2214 CTTCT..................GGCTCCCTGTATAAGTCCAGACTGAAACC 2180
     |::: ||| |||||
  39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55

2179 CCCTTGGAAGGCCTCCAGTCAGGCAGCCCCTAGAGACTGGGGAGGAGGAG 2130
     ||| |||||||| ||||||||
  56 Pro...GlnGlyLeu...ProGlySerPro............GlyAlaPr 66

893 TGAGCAGGCCAAAGAGGCACTCCT...............CCTGGGTGCCC 859
    ||| ||| ||||| || ||
494 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 510

858 AGGTAGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA..... 813
    |||:::||| ||| :: |||||||
510 oGlyGlyGlnGlnHisHisLeuGlyLysGlnAlaGlyGlyAspV 527

812 ..GGTAGCCCAGGCAGCCCCAAAGACTGATCATGAAGGCATAGACAGAGT 765
    ||||||||||||| ||| ::||| |
527 aIGlySerProGlyAlaPro.............Gly 534

764 AGGCCTGGCGACAGTGT...CCGGGT..CCCGGAAGAGGTCAGAGAGC 721
    |||:::|| |||||| ||| ||||||
535 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 551

720 AGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGAGCCC 671
    ||||| ::|||:||||| ||::|::::|
551 oGlyProGlnGlyLeuProGlySerProGly.....AlaProGlyThrP 566

670 CACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATCCGGGCACA 621
    ||||| |||||||| ||||||||
566 roGlyProGlnGlyLeuProGlySerProGlyAla.......ProGlyThr 582

620 GCAGCCCTGCTAGCCAGCCGCCCCTTGGGATGATGAGAAAGA....... 582
    ::::   ::|||
581 ProGlyGluGlyGlnGlnHisHisLeuGlyLysGlyAlaLysGlnAlaGlyAs 597

581 ....GGCTCAGCAGGATGCCCAAGGACAGTG...CCCAGATGAAGGGCCG 539
    ||||||| ::||||| :::|||
597 pValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG 614

538 GCGGGCGGCCATAGCCTCCACGCCAGTGGTCACTGGCTGCCTAGGAGCG 489
    |||||| ||| ||||||
614 ly.........SerProGlyAla 618

488 GGACAGAGACCAGGCCCAGCACTGGACCAATGCCCAGCACCATGGTCATG 439
    ||||||||| |||
619 ProGlyThrProGlyPro 624

438 AACTTCTCCTCTACCCCCACTTCCAGCAGCAGAGGCGGCACATAGGTGAT 389
    ::: |||||| |||::::||
625 ...GlnGlyLeuProGlySerProGlyAla 633

388 GCCTCGCGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCA 339
    ||| ||| ||| ||| ::::|||
634 ....ProGlyThrProGlyProGlnGlyLeuProGlySerProGly 647

338 AGAGCTGGCTTTCCGGT............................ 321
    ||| |||
648 AlaProGlyThrProGlyGluGlnGlnHisHisLeuGlyGlyAlaLy 664

320 ....GCCCAGCAGGCGGCTCACCCACCCACCAGCCTCTGGACCATAGTGGCCA 275
    |||::: |||||||| |||
664 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyPro 680

274 GGCGGG...TAGGGGGGCCGGTTCAGGCGGCCACTCCAGAACTGCTTCG 228
    ||| |||||||| |||
681 GlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGl 697

227 TCTCGGCTGTCTGCCCAGAAGCTCGCGGCCTTCCTTCCTTGCTGCCGCCAAC 178
    ||:::: ||| |||||||| :::|||
697 yLeuProGlySerProGlyAlaProGlyThrProGly.............. 709

177 TGCCTAGGAATCAGGCAGGCGGCCCATTTCTGCCAGCCCCTTTGGTGCCGGT 128
    |||:::::: ||||||||::   :: ||| |||
710 ..ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrProGly 724

127 CCAGCTTCTCAGCCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCAGTG 78
    :::|||:::|| ||||||
725 GluGlyGlnGlnHisHisLeuGly........GlyAlaLysGlnAl 737

77 GGGACAACGTCTCATCA........CTCAGAATCCTGGCCGA 46

Wed Sep 29 14:26:53 1999

2129 AGGGACGCCCCAGCCCCCAGTCTGTGCGAGCTACGCACCTCAGCA......G 2086
     ||||||||| ||| :::::: ||| ::|||
  66 oGlyThrProGlnGlyLeuProGlyLeuProGlySerProGlyAlaProGlyT 83

2085 CACAGGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAA.......... 2046
     :: ||||||::::::   :: |||||
  83 hrProGlyGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99

2045 ..........CAGAAACTGGGCGCCCAGCCCGCGAGCCCCATGGGG 2011
              ||| ||||||::::||::|||||
 100 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 116

2010 CTAACAGGAGGGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCACCCAA 1961
     ::|||| ||| |||::||:|||||
 116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro. 132

1960 TGTGCTGGAAGTTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTCAA 1911
                                        ::: |||
 133 ....GlyProGln 135

1910 ATACTACCTGTGTAGCAAAGTAAATGGGACCAGACCCAGGCCTGCCGGCA 1861
                                ::|||||
 136 GlyLeuPro...........................GlySerProGlyAlaProGlyl 145

1860 GACACCATATAGGCAGTGACAGACTGGCTGAGCTGACAATGGAGCCCAT 1811
     |||||| |||:::|||| |||||
 145 yThrProGlyGlyAlaLysGlnAlaGlyAspValGly.......SerPro. 159

1810 AAACAGGGATGGGCCACCTGGGACACAGGAAGGCAC........... 1773
              ||| |||||||| ||| |||::
 160 ....GlyAlaProGlyThrProGlyProGlnGlyLeuProGly 172

1772 ........TATCCAGATGGCGAGGTCCAGGCAGATGCCCGGC..... 1737
             ||||| |||||| ::|||||
 173 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 189

1736 .......CCGGAACCACCCTGGCCTCCGTG......GGCTCACCCACCA 1701
            ||| ||| :::|| ||||||||| :
 189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA 206

1700 CCACACGTACG..........GAGACATCACAGGCAGAGGCCCG 1666
     :: ||||::::   :::|||
 206 laProGlyThrProGlyGlyAlaLysGlnAlaGlyAspValGlySerPro 222

1665 CAGAGGCCGGGGGTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCACCGTG 1616
     :: ||| :::|| ||||||||| :::|||
 223 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 239

1615 TCCATTAGGGAAGGGAGCTCCAGGCTTAGGGCCTGGCAGGAAGCTGGTCA 1566
     :||| |||||| :::|||
 239 aProGlyThrProGlyProGlnGlyLeu...ProGly.SerProGlyAla 254

1565 TCAGGCTGTCCTCACTGCTAGCACCTCCAGTGTCCCCTCGGTATTTGGGC 1516
                                ||| ::::
 255 ProGlyThrProGly.......................ProGlnGlyLeu..... 263

1515 AGGAACACCTGCTTCTCCCGGTGGTAGAGGAGGCCA......GTGTGTA 1472
     ||::::||||||||::: ||| ::: :
 264 ....ProGlySerProGlyThrProGlyGlyAlaLysG 278

1471 GGGCAGGATCTGCAGGGGATGGAGATGAGAAAGAGGCTCAGCAGGACCCACA 1422
     ::||||||::::::||||| ::: |||
 278 lnAlaGlyAspValGly.............................. 283

1421 CTGTCACCACGGCCACACTGTGGACAGGCATGTGGCACCGGCCAGCCACA 1372
     ||||||::|||||| ::: ||||
 284 ..SerProGlyAlaProGlyThrProGly...........ProGl 294

1371 GGGAAAGCTGCCACCACTGGCCAAATAGACTGCTCGAGTGCCGAATCCGTG 1322
     |||| |||||| ::::::
 294 nGly...LeuProGlySerPro.........................GlyA 302

1321 CACCAGCCGGTCCATGACCAGCAGAGAGAAGAACCAGGGAGATGGGCCACTGCA 1272
     ||||||:::: ||| |||||
 302 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 318

1271 GGAACA.............................GCCCCAGGCTG 1255
     ||||||
 319 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335

1254 CCCA...TCCGAACGCCTTCATCATAGTGTCTCCGGGCCTCGGTGCCCGG 1208
     ||| ::: || |||||
 335 rProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG 352

1207 CTCAGCTCTGGGCACGC...........CCTGGTACAGGCCCCTCGCCCACGAAAT 1164
     ::: |||||
 352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368

1163 CCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGC.......... 1125
     :::   ::||::::|||
 369 ProGlyProGlnGly.....................LeuProGlySerProGlyAlaPr 381

1124 .TGCACAGCTCAGCCACCAAGAGCCGGCCAGGGTGCCGGGGCATGCGGCA 1076
                 ||| ::: || ||
 381 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 398

1075 GCACAGCTGGTGCA...........GCCGGGGAAGCAGGGGCGCCCAGGTCC... 1035
     :: |||||| |||:::|||||
 398 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 414

1034 GGAAAGCCAAGCGGCCCCGCATGGACAGCAGTGGGCGCCACAAGGAGGGG 985
     ||| :::   ::: ::|||
 415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431

984 GCCGACAGCCCTTCTGCTGGCTGGGCCCA...GCGCTGCCTCCTC 938
    ::: ||| :::|||||
431 rPro.............................GlyProGlnGlyLeuProGlyLys 440

937 AGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCA 888
    ::||| ||
440 erPro.....................GlyAlaProGlyThrPro 447

887 GGGCCAAAGAGGCACTCCT......CCTGGGTGCCCAGGTAGGGGCCAGG 844
    ||| ||| :::|||||
448 GlyProGlnGlyLeuProGlyLeuProGlyThr...ProGl 463

843 GCACTGGTGTCCCAGTCAATGGCAGGCAGGA......GGTAGCCCAGGCA 800
    | :::|||||
463 y.........GlyAlaLysGlnAlaGlyAspValGlySerProGlyA 476

799 GCCCCCAAGACTGATCATGAAGGCAGCATAGACAGTAGGCCTGGCGACAGT 750
    ::|| ::: |||
476 laPro...................GlyThrProGlyProGln 483

749 GGT...CCGGT...CCCGGAAGAGGTCAGAGAGCAGGGCCTCCAGTGA 706
    ||| |||||
484 GlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe 500

705 GTGAAGCACACCTGGCCACAGAGTCCAGCCGCCACGCCCAGGATGAG 656
    ::: ||||||
500 uProGlySerProGly.....AlaProGlyThrProGlyProGlnGlyLyL 515

655 CAGTGCCAGCTCCAGGGGCCTGGGATCCGGGCACAGCAGCCCTTGCTAGCC 606
    :::|||||||| ||||||||||
515 euProGlySerProGlyAla................ProGlyThrProGly......Gly 527

605 AGCCGGCCCTTGGGATGGATGAGAAAGAGGCTCAGCAGGATGCCAAGGACAGT 556
    ::::: ||||||| ||| ::::: |||
528 AlaLysGlnAlaGly..AspValGlySerProGlyAlaProGlyThrPr 543

555 G...CCCAGATGAAGGGCCGGCGGCGGCCATAGCGTCCACGCCAGTGGTC 509
    ||::: ||
543 oGlyProGlnGlyLeuProGly.......................GlyA 550

508 ACTGGCTGAGCCTAGGAGCGGGACACAGACACCAGCCCACCACTGGACCAA 459

Wed Sep 29 14:26:53 1999

; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 682 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-397-633A-36

alignment_scores:
                    Quality: 370.50   Length: 831
                    Ratio: 1.015   Gaps: 59
Percent Similarity: 43.923   Percent Identity: 29.964

alignment_block:
US-09-030-606-110/rev x US-08-397-633A-36

Align seg 1/1 to: US-08-397-633A-36   from: 1 to: 682

2314 AAAAACCCTTCTCTAGGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     ::||||||| ||||||:::||||||:::
  12 GluAsnPro.....GlyValThrGlnLeuAsn..........ArgLeu 22

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCCGCCATTCCAGTGCATGGAGCC 2215
     |::: ||| ||||||:::|||| |||:::|
  22 uAlaAlaHisProProPheAlaSerAspProMetGlyAlaProGlyThrP 39

2214 CTTCT........GGCCTCCCTGTATAAGTCCAGACTGAAACC 2180
     ||::: ||| |||
  39 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 55

2179 CCCTTGGAAGGCCTCCAGTCAGCAGCCTAGAGACTGGGGAGAGGAGGAG 2130
     |||||||||| ||||||||
  56 Pro...GlnGlyLeu...ProGlySerPro..........GlyAlaPr 66

2129 AGGGACGCCCAGCCCCCAGCGTGTGCAGCTACGCACCTCAGCA......G 2086
     ||||||| :::||:   ::|||
  66 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 83

2085 CACAGGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAA.......... 2046
     :: |||||:::|||::::
  83 hrProGlyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro 99

2045 ...........CAGAAACTGGGGCCCAGCCCGCCAGCCCCATGGGG 2011
     :::|||||||| ||| ::||||||:::|||||||||
 100 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 116

2010 CTAACAGGAGGCGGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCACCCAA 1961
     :::||| ||| |||||||||||:::||||||
 116 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro. 132

1960 TGTGCTGGAAGTTTTCTACGCTGCTGATATTTGGCCAAGTCGCTTCTGTCAA 1911
     ||| ||||||||| ||| ::: |||
 133 ....GlyProGln 135

1910 ATACTACCTGTGTAGCAAAGTAAAATGGGACCCAGGCCTGCGGCA 1861
     :::||||||
 136 GlyLeuPro.................GlySerProGlyAlaProGl 145

1860 GACACCATATAGGCAGTGACGACTGGCTGAGCTGGACAATGGAGCCCAT 1811
     |||:::|||||| ::|||| :::||| ||||||
 145 yThrProGlyGlyAlaLysGlnAlaGlyAspValGly......SerPro. 159

1810 AAACAGGGATGGGGCCACCTGGGACAGCAGGAGGCAC............. 1773
     ||| |||||||||| ||| :::
 160 ....GlyAlaProGlyThrProGlyProGlnGlyLeuProGly 172

1772 .........TATCCAGGATGGCGAGTCCAGGCAGATGCCCCGGC....... 1737
     ||||| |||||
 173 SerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerPr 189

1736 ......CCGGAACCACCCTGGCCTCCGTG......GGCTCACCCACCA 1701

───

551 ....SerProGlyAlaProGlyThrProGlyPro............. 560
         ||| ||||||

458 TGCCCAGCACCATGGTCATGAACTTCTCCTACCCCACTTCCAGCAGC 409
                                          ||||:::|
561 ....GlnGlyLeuProGlySerProGlyAl 569
                   :: ||||||

408 AGAGGCGGCACATAGGTGATGCCTGCGGGCAAACACACCTCCAGGCCAAA 359
                  |||:::|||||| |||
569 a................ProGlyThrProGlyProGlnG 577
                      |||:::||||||| |||

358 GGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCCGGTGCCGCA...... 315
     ||| ||| ::: |||:::|| |||:::|||
577 lyLeuProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysGln 593

314 GCAGGC.......GGCTCACCCACAGCCCTGGACCCATAGTGGGCCAGGCG 271
     |||||| |||||||||
594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610

270 GG...TAGGGCTCAGGGGCGTTCAGGCACTCCAGAACTGCTTCGTCTC 224
     || |||||| ||| |||||| |||||||
610 lyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeu 626

223 GGCTCTGCTCCAGAAGTCGCGGCCTCTCCTCCTTGCTGCCGCCAACTGCC 174
     ::::||| ||| :::|||||| ||||||||||
627 ProGlySerProGlyAlaProGlyThrProGly................Pr 638

173 TAGGAAATCAGCCAGGCGCCCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAG 124
     |::::::: ||||||:|| ||| ::: |||
638 oGlnGlyLeuProGlySer.............ProGlyAlaProGlyThrP 651

123 CTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCCACCTCAGTGGGGA 74
     ||::: ||| ||| |||
651 roGly.............................GlyAlaLysGlnAlaGly 658

73 CACGTCTCATCA..........CTCAGATCCTGGCCGA 46
   ||||||||||| :::|||||||||||
659 AspValGlySerProGlyAlaMet.AspProGlyArg 670

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-36

seq_documentation_block:
; Sequence 36, Application US/08397633A
; Patent No. 5773577
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATESCAPABLE
; TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
; NUMBER OF SEQUENCES: 105
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/397,633A
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/397,633A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299

```
 189 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyA  206
          |||           |||   ||||::
1700 CCACCGTACG..........GAGACAATCACAGGAGAGAGGCCCG      1666
 206 ..laProGlyThrProGlyAlaLysGlnAlaGlyAspValGlySerPro  222
      ::  :|||            ||||||| :::|||
1665 CAGAGCGCGGTGGAGTGGGAGCAGGCCACTGCCTTCCAGCACCCACGTG  1616
 223 GlyAlaProGlyThrProGlyGlnGlyLeuProGlySerProGlyAl    239
     ::: |||   ||| |||||||
1615 TCCATTAGGGAAGGAGCTCCAGGTTAGGCCTGGCAGGAAGCTGGTCA    1566
 239 aProGlyThrProGlyProGlnGlyLeu...ProGly.SerProGlyAla 254
      |||          |||||| ||||||
1565 TCAGGCTGTCCTCACTGCTACCACCTCCAGTGTCCCTCGGTATTTGGGC  1516
 255 ProGlnGlyLeuProGlyThrProGly..........ProGlnGlyLeu.....  263
     ||| |||         |||
1515 AGGAACACCTGCTTCTCCCGTGGTAGAGGGAGGCCA.....GTGTGTA   1472
 264 .......ProGlySerProGlyAlaProGlyThrProGlyGlyAlaLysG 278
            :::|||||||:: ||| |||
1471 GGGCAGGATCTGCAGGGCTGAGAGGTGAACCCGGTGAGGGCGGCTGAAG  1422
 278 lnAlaGlyAspValGly.......                            283
     ::|||||||:::|||
1421 CTGTCACCACGGCCACACTGTGGGACACAGGCATGTGCACCGGCCAGCCACA 1372
 284 ...SerProGlyAlaProGlyThrProGly..........            294
        ||||
1371 GGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTG 1322
 294 nGly..LeuProGlySerPro.........GlyA                  302
     |||   ||||||                 :::|
1321 CACCAGCCGGTCCATGACCAGAGAAGACCAGGAGATGGCGACACTGCA   1272
 302 laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 318
      :: |||::|| |||||:|:::|||
1271 GGGAACA.....GCCCCAGGCTG                             1255
 319 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 335
     |||||| |||
1254 CACCAGGCGGCTTCATCATAGTGTCTCCGGGCCTCGGTGCCGG        1208
 335 rProGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProG    352
      ::|||::  |||   :::|||
1207 CTCAGCTCTGGCACGC.....CCTGGTACAGCCCCTGCCCCACGAAAT   1164
 352 lyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr 368
      |||||| ||| :::
1163 CCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGC..........  1125
 368 ...........LeuProGlySerProGlyAlaPr                 381
                :: |||||
1124 .TGCACAGCTCAGCCACGAAGACGCCGGCCAGGGTGCGGGGCATGCGGCA 1076
 381 oGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyT 398
      |||   :::|||||| |||::::|||
1075 GCAAGCTGTGCA.....GCCGGGGAAGCAGGGCGCCCAGGTCC...     1035
 398 hrProGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaPro    414
      :::||||:| :::|||
1034 GGAAAGCCAAGCAGCGGGCCGCCATGGACCAGCAGTGGGCAAGGAGGGGG 985
 415 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 431
     ||| ||: :::|||
 984 GCCCAGCAGCCCTTCTGCTGGCTCGGTGGGGCCCA....GCGCTGCCTCCTC 938
 431 rPro...........GlyProGlnGlyLeuProGlyGlyLyS          440
     :::                :::|||||||
 937 AGCCACCAGCAGTGTGGCTGCTACCGGAGTGAGAAGATGAGGGTGAGCA  888
     ::|||                              |||
```

```
 440 erPro.................................GlyAlaProGlyThrPro 447
                                          ||||||
 887 GGCCAAAGAGGCACTCCT.....CCTGGGTGCCCAGGTAGGGGCCAGG   844
 448 GlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProGl 463
     ::::||||                ::||||:||:|||
 843 GCACTGTGTCCCAGTCAATGGCAGGCAGGA.....GGTAGCCAGGCA    800
 463 y...........GlyAlaLysGlnAlaGlyAspValGlySerProGlyA  476
     |                ||||||| :|||
 799 GCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGCCTGGCGACAGT  750
 476 laPro............................GlyThrProGlyProGln 483
     :::  |||||| :::
 749 GGT...CCGGT...CCCGGAAGAGGTCAGAGAGCAGGGCCTCCAGTGGA  706
 484 GlyLeuProGlySerProGlyAlaProGlyThrProGlnGlyLe       500
     ||||||                  |||| :
 705 GTGAAGCACACTGGCCACAGAGTCCAGCCCACGCCCAGGATGAG       656
 500 uProGlySerProGly.....AlaProGlyThrProGlnGlyL        515
     |||:||||||          ::  :|||
 655 CAGTGCCAGCTCCAGGGCCTGGATCCGGCACAGCCCCTGCTAGCC      606
 515 euProGlySerProGlyAla.....ProGlyThrProGly.....Gly   527
     ::::|||||||          |||          ::::
 605 AGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGCAGATGCCCAAGGACAGT 556
 528 AlaLysGlnAlaGly...AspValGlySerProGlyAlaProGlyThrPr 543
     || :::||||       :: :||||||  |||::::|| |||
 555 G...CCCAGATGAAGGGCCGGCGGCCATAGCGTCCACGCCAGTGGTC    509
 543 oGlyProGlnGlyLeuProGly..............               550
     |||
 508 ACTGGCTGAGCCTAGGAGCGGGACACAGACCCAGCACTGGACCAA      459
 551 .........SerProGlyAlaProGlyThrProGlyPro..........  560
               ||| |||::::||||||
 458 TGCCCAGCACCATGGTCATGAACTTCCCTACCCCCACTTCCAGCAGC    409
 561 .............GlnGlyLeuProGlySerProGlyAl            569
                  :::|||||| :||||||
 408 AGAGGGCCGCACATAGGTGATGCCTCGGCCCAAACACACTCCAGGCCAAA 359
 569 a...........................ProGlyThrProGlyProGlnG 577
     |                            ||| :::|||
 358 GGTTAGCAGGTTGACCAGCAAGACGTGGCCTTCCGTGGCCGCA.....   315
 577 lyLeuProGlySerProGlyAlaProGlyThrProGlyAlaLysGln    593
      ||| :::||||:||||||
 314 GCAGGC.....GGCTCACCCACCAGCCTCTGGACCATAGTGGGCCAGGCG 271
 594 AlaGlyAspValGlySerProGlyAlaProGlyThr.ProGlyProGlnG 610
     ||||||||| :: |||||| |||        ::: |||
 270 GG...TAGGGGTCAGGGGCCGTTCAGGCACTCCAGAACTGCTTCGTCTC  224
 610 lyLeuProGlySerProGlyAlaProGlyThrProGlnGlyLeu       626
      |||       :::|||||||:::|||
 223 GGCTCGTCTGCCAGAAGCTGCGGCCTCCTCCTTGCTGCCGCCCAACTGCC 174
 627 ProGlySerProGlyAlaProGlyThrProGly.............Pr   638
     ||| :::|||||||                   ::
 173 TAGGAATCAGCCAGGCCGCCCATTTCTGCCCAGCCCTTTGTGCCGGTCCAG 124
 638 oGlnGlyLeuProGlySer..........ProGlyAlaProGlyThrP   651
      :::|||||||
 123 CTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGCACCTCAGTGGGA   74
 651 roGly................................GlyAlaLysGlnAlaGly 658
      |||                              :::||||||
  73 CACGTCTCATCA.........CTCAGATCCTGGCCGA              46
     :::|||
 659 AspValGlySerProGlyAlaMet.AspProGlyArg             670
```

us-09-030-606-110.rai

Wed Sep 29 14:26:53 1999

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-707-237A-84

seq_documentation_block:
; Sequence 84, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
;   APPLICANT: Ferrari, Franco A.
;   APPLICANT: Capello, Joseph
;   APPLICANT: Crissman, John W.
;   APPLICANT: Dorman, Mary A.
;   TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
;   TITLE OF INVENTION: REPETITIVE DNA
;   NUMBER OF SEQUENCES: 108
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: United States
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/707,237A
;   FILING DATE: 03-SEP-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/175,155
;   FILING DATE: 29-DEC-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/053,049
;   FILING DATE: 22-APR-1993
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/609,716
;   FILING DATE: 06-NOV-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/269,429
;   FILING DATE: 09-NOV-1988
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/114,618
;   FILING DATE: 29-OCT-1987
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 06/927,258
;   FILING DATE: 04-NOV-1986
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Trecartin, Richard F.
;   REGISTRATION NUMBER: 31,801
;   REFERENCE/DOCKET NUMBER: A-55186-10/WHD
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 84:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 761 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: protein
US-08-707-237A-84

alignment_scores:
    Quality: 365.00   Length: 751
    Ratio: 1.058   Gaps: 52
Percent Similarity: 45.939   Percent Identity: 30.093

alignment_block:
US-09-030-606-110 x US-08-707-237A-84

Align seg 1/1 to: US-08-707-237A-84   from: 1 to: 761

84 GGTGCCCACAGCAGCAGGTGTTG......AGCATGGCTGAGAAGCTGG 127
   |||::||| ||| || :|| ::
51 GlyThrProGlnGlyLeuProGlySerProGlyAlaProGlyTh 67

128 ACCGGCACCAAAGGC.......TGGCAGAAAATGGGCGGCCGC.... 164
    |||::||| ||| ::: ||||||||
67 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 84

165 .........TGATTCCTAGGCAGTTGGCGGCCAGCAAGGAGGAGAG 200
         |||||::: ::||||||: :||
84 lyGluGlyGlnGlnHisHisLeuGlyGly..AlaLysGlnAlaGlyAsp 99

201 GCCGCAGCTTCTGGACGACGACCGAGACGAAGCAGTTCTGGAGTGCCTGA 250
    ::::::||: ::||: :: ::|||||||| ||
100 valGlySerProGlyAlaProGlyThrProGlnGlyLeuProGlnGlyLeuProGl 116

251 ACGGCCCCTGAGCCCTACCCGCCTGGCCCACTATGGT........... 288
    ||| ||| ||||||||| |||
116 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 132

289 CCAGAGGCTGTGGGTGAGCCG...CCTGCTGCGGCACCGGAAAGCCCAGC 335
    ||| ||| ||| ||| ||: |||::||||||| ||||||
133 ProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGl 149

336 TCTTGCTGGTCAACCTGCTAACCTTGGCCTGGAGGTGTGTTTGGCCGCA 385
    : ||||||||| |||::
149 yAlaProGlyThrProGlyGlu................GlyGlnG 159

386 GGCATCACCTATGTGCCGGCCTCTGCTGCTGGAAGTGGGGTAGAGGAGAA 435
    ::||||||||| ::|||::||: ||||||:::|| |||
159 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGly.......... 172

436 GTTCATGACCATGGTGCTGGCCATTGGTCCAGTGCTGGCCTGGTCTGTG 485
    ||| |::|||||||
173 ....SerProGlyAlaProGly...Th 179

486 TCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGC 535
    ||||::||::|| ||
179 rProGlyProGln................ 183

536 CGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCT... 582
    :: :: ::: |||::: |||
184 ....GlyLeuProGlySerProGlyAlaProGly 193

583 CTTTCTCATCCAAGGGCCGGCTGGTCTAGCAGGGCTGCT....GTGCC 626
    |||::||| ::||:||
194 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 210

627 CGGATCCCAG..........GCCCCTGGAGCT...GGCACTGCTCATC 661
    ||| || ::||||||||| |||||::||
210 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.......P 225

662 CTGGCGTGGG..............GCTGCTGGACTTCTGTGGCCAGTGTG 699
    ::||| :::|||
225 roGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAsp 241

700 CTTCACTCCACTGGAGGCCCTGCTCTCTGACCCTTCCGGGACCC...... 744
    ||| :: ||||||||
242 valGlySerProGlyAlaProGly........ThrProGlyProGlnGl 255

745 ..................GGACCACTGTCGCCAGGCCTACTCTGTCTAT 775
          ::|||| |::||||||||
255 yLeuProGlySerProGlyAlaProGlyThrProGly............. 267

776 GCCTTCATGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCTGCCAT 825
    ||||||||| ||| |||
268 ....ProGlnGlyLeuProGlySerProGlyAla...... 277

826 TGACTGGGACACCAGTGCCCTGGCCCC.........CTACCTGGGCACCCAGG 869
    |||||||||| |||

Wed Sep 29 14:26:53 1999

```
278  ........ProGlyThrProGlyProGlnGlyLeuProGlySerPro..  290
870  AGGAGTGCCTCTTTGGCCTGCTCACCCTACCTTCCTCACCTGCTAGCA   919
           ||||||:|||     :::||    |||||||         :::::
291  .GlyAlaProGlyThrProGlyGluGlyGlnHisHisLeuGlyGlyA   307
920  GCCACACTGCTGGTGGCTGAGGAGGCAGCCGCTGGGCCACCGAGCCAGC  969
       ::   |||||          |||::   ||||||     :::::|||
307  laLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly  323
970  AGAAGGGCTGTCGGCCCCCTCCTTGTGCGCCCACTGCTGTCCATGCCGGG 1019
       :::::::||||||        :::    ||||||    GlyThrProGl
324  ProGlnGlyLeuProGlySerProGlyAlaPro......GlyThrProGl  338
1020 CCCGTTGGCTTTCG...GAACCTGGGCGCCCTGCTTCCCGGCTGCAC    1066
       |||    ||||||     ||||:|::::|||    :::|||
338  yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP  354
1067 CAGCTGTGCTGCCGCCATGCCCGCCACCCTGCGCCGCCTTCGTGGCTGA  1116
       ||   |||||      ::|||    ||||||::::||:||
354  roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu  369
1117 GCTGTGCAG.......CTGGATGGCACTCATGACCTTCACGCTGTT....  1155
       ::|   ||| |||   ::   ::     ||
370  GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe  386
1156 ......TTACACGGATTCGTGGGCGAGGGCTGTACCAGGGGCGTGCCC   1198
           :::||     ||||          :::::|:|||
386  rProGlyAlaProGlyThrProGlyLeuProGlySerProGlyAlaProG  403
1199 AGAGCTGAGCCGGGCACCGAGCCCGGAGACACATGATGAAGGCGGTTCG  1248
       ::::      ||||||       :::|||   ::::||||||
403  lyAla...ProGlyThrProGlyProGlnGlyLeu..............  413
1249 GATGGGCAGCCTGGGGCT......GTTCCTGCAGTGCGCCATCTCCCTGG 1292
       |||::|:||||||||      :::|||||    ::   |||||||
414  ProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGl  430
1293 TCTTCTCTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCAC.......  1335
       |     :::||     ||:   ::||||||     :::||
430  ySerProGlyAlaProGlyThrProGlyGluGlyGlnHisHisLeu     447
1336 ..TCGAGCAGTCTATTGGCCAGTGTGGCAGCTT...TCCCTGTTGGCTG  1379
         :::       |||   ||||:::|||   ||||||
447  lyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThr  463
1380 CCGGTGCCACATGCCTGTCC................ACAGTGTG       1408
       ||| |||:::||    |||                   ||
464  ProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGl  480
1409 GCCGTGGTGACAGCTTCAGCCGCCCCTCACCGGGTCACCTTCTCAGCCCT 1458
       |||     |||   ||||||||||:::||||||||::::||
480  yPro.......GlnGlyLeuProGlySerProGlyAlaProGly.ThrPro  494
1459 GCAGATCCTGCCCTACACACTGGCCTCCCTTCTACCACGGGAGAAGCAGG 1508
       |||||||:||||||     |||||      :::||
495  GlyPro.........GlnGlyLeuProGlySerProGlyAlaProGl     507
1509 TGTTCCT................GCCCAAATACCGA               1528
       |:::|||                  || ||
507  yThrProGlyGluGlyGlnHisHisLeuGlyGlyAla.LysGlnAla    523
1529 GGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCT 1578
       ||||||:||:::    :::::       :|
524  GlyAspValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLe  540
1579 GCCCAGGCCCT.....AAGCCTGGAGCTCCCTTCCCTAATGGACACGTGG 1622
       |||||||::||     |||:::|||||||:::|
540  uProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProG  557
1623 GTGCTGGAGGCAGTGGCCTGCTCCACCTCCACCCGCTCTGCGGGGGCC   1672
       ||:::: |||||     :::||||  ||||||::::|||
557  lySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySer  573
```

```
     ...............................................  1672
574  ProGlyAlaProGlyThrProGlyGluGlyGlnHisHisLeuGlyGl    590
1673 ......TCTGCCTGTGATGTCTCCGTACGTGTGGTGGTGAGCCCA     1715
           :::|||    ||              |||||  ||||||
590  yAlaLysGlnAlaGlyAsp..............ValGly.SerPro    600
1716 CCGAGGCCAGGTGGTTCGGGCCGGGCATCGGCTGA.........    1755
       ||| |||     |||||    ||||||
601  GlyAlaPro.GlyThrProGlyLeuProGlySerProGlyA        617
1756 ..CCTCGCCATCCTGA...TAGTGCCTTCCTGCTCCTCCAGGTGGCCCC 1800
       |||    ||||||     :::::   |||||    ::::::|||
617  laProGlyThrProGlyLeuProGlyLeuProGlySerProGlyAlaPro 633
1801 ATCCCTGTTTATGGCCTCCATTGTCCAGCTCAGCCAGCTCTGTCACTGCCT 1850
       ||||||     :::||:||
634  GlyThrProGlyProGlnGlyLeuProGlySerPro..........    645
1851 ATATGGTGCTGCCGCAGGCCTGGGTCTGGTCGC...CATTTACTTT    1894
       ||::|     |||   ||||||       |||||    |||||||
646  ....GlyAlaProGlyThrProGlyGluGlyGlnHisHisLeu.....   659
1895 GCTACACAGGTAGTATTTGACAGAGCGACTTGGCCAAATACTCAGCGTA  1944
       ||:::     :::       ||| |||:::||
660  .........GlyGly.....AlaLysGlnAlaGlyAspValGlySerPro 671
1945 GAAAACTTCCAGCACATTGGGTGAGGGCCTGCCTCACTGGGTCCCAGC   1994
       ||     :::||||||||         :::|||||||
672  GlyAlaProGlyThrProGlyLeuProGlySerProGlyAl        688
1995 TCCCCGCTCCTGTTAGCCCCATGGGGCTGCCGGGCTGGCCG          2035
       | ::::    :::||:::::::||||||:::||||||
688  a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro         701
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-114

us-09-030-606-110.rai

Wed Sep 29 14:26:53 1999

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  762 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
; US-08-642-255-114

alignment_scores:
        Quality:   365.00              Length:      751
          Ratio:     1.058               Gaps:        52
Percent Similarity:  45.939      Percent Identity:  30.093

alignment_block:
US-09-030-606-110 x US-08-642-255-114

Align seg 1/1   to: US-08-642-255-114   from: 1   to: 762
```

```
 226 roGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAsp 242

 700 CTTCACTCCACTGGAGGCCCCTGCTCTCTGACCTCTTCCGGGACCC...... 744
     ||||:|||     |||::||:::||           |||||||||
 243 valGlySerProGlyAlaProGly.........ThrProGlyProGlnGl 256

 745 .................GGACCACTGTCGCCAGGCCACTACTCTGTCTAT 775
                      ||||:|||::|||
 256 yLeuProGlySerProGlyAlaProGlyThrProGly.............  268

 776 GCCTTCATGATCAGTCTTGGGGGTGCCTGGGCTACCTCCTGCCTGCCAT 825
     ||||||||||||        ||        |||
 269 .........ProGlnGlyLeuProGlySerProGlyAla..........  278

 826 TGACTGGGACACCAGTGCCCTGCCCC.......CTACCTGGCACCCAGG 869
                                ||||||||    ||||||||
 279 ..........ProGlyThrProGlyProGlnGlyLeuProGlySerPro. 291

 870 AGAGTGCCTCTTTGGCCTGCTCACCCTACATTCCTCACCTGCGTAGCA 919
     :::|::|||        |||    :::         ::
 292 .GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyA 308

 920 GCCACACTGCTGGTGCTGAGGAGGCAGCAGGCCCCCACCGAGCCAGC 969
     ||||:|              ::||||:||:::
 308 laLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324

 970 AGAAGGGCTGTCGGCCCCTCCTTGTCGCCCACTGCTGTGTCCATGCCGG 1019
     :::||:::                |:::|||
 325 ProGlnGlyLeuProGlySerProGlyAlaPro.......GlyThrProGl 339

1020 CCCGCTTGGCTTCCG...GAACCTGGGCGCCGCCTTCCCCGTGCTGCAC 1066
     ||| ||:::|||    :::||:||||   |||:|||          :::|
1067 CAGCTGTGCTGCGCATGCCCCGCACCCTGCGCCGGCTTCGTGGGCTGA 1116
     || |||||||                         |||||::::
 355 roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 370

1117 GCTGTGCAG........CTGGATGGCACTCATGACCTTCACCGTGTT..... 1155
     |||||::|                  |||  ||||||
 371 GlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 387

1156 .......TTACACGGATTTCGTGGGCGAGGGGGCTGTACCAGGGCGTGCCC 1198
                              ||| :::||||::|||||
 387 rProGlyAlaProGlyThrProGlyLeuProGlyLeuProGlySerProG 404

1199 AGAGCTGAGCGGGCACCGAGCCCCGGAGACACTATGATGAAGGCGTTCG 1248
     :::|            |||      ::|||
 404 lyAla...ProGlyThrProGlyProGlnGlyLeu.............. 414

1249 GATGGGCAGCCTGGGGCT........GTTCCTGCAGTGCGCCATCTCCCTGG 1292
     :|||||||||||              ::|:|||||   ||||||||||
 415 ProGlySerProGlyAlaProGlyThrProGlnGlyLeuProGl 431

1293 TCTTCTCTGGTCATGGACCGGCTGGTGCAGCGATTCGGCAC........ 1335
     :|:::||          |||    ::|||
 431 ySerProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuG 448

1336 .TCGAGCAGTCTATTTGGCCAGTGTGGCAGCTT....TCCCTGTGGCTG 1379
     ||          |||  ||||||          ||||||
 448 lyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464

1380 CCGGTGCCACATGCCGTCCC................ACAGTGTG 1408
     ||| |||:::|||
 465 ProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGl 481

1409 GCCGTGGTGACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCT 1458
     |||       ||||||         ||||||||:::|||
 481 yPro.....GlnGlyLeuProGlySerProGlyAlaProGly.ThrPro 495

1459 GCAGATCCTGCCCTACACACTGGCCTCCTACACCGGGAGAAGCAGCAGG 1508
                    ||||||            ||||||        |||
 496 GlyPro.......GlnGlyLeuProGlySerProGlyAlaProGlyGl 508
```

```
  84 GGTGCCCCAGCAGCAGGTGTTG......AGCATGGGCTGAGAAGCTGG 127
     ||:::|||     |||   |||          :::
  52 GlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyTh 68

 128 ACCGGCACCAAAAGGGC......TGGCAGAAATGGGCGCCTGGC....... 164
     ||:::||||:::|||           |||||||
  68 rProGlyGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 85

 165 ........TGATTCCTAGGCAGTCAGTTGCCGGCAGCAAGGAGGAGAG 200
                 ||||||||          :::||||||:::
  85 lyGluGlyGlnGlnHisHisLeuGlyGly...AlaLysGlnAlaGlyAsp 100

 201 GCCGCCAGCTTCTGGAGCAGCAGGCCGAGCAAGCAGTTCTGGAGTGCCTGA 250
     :::::::|       ||||||         ||  ||||||
 101 valGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGl 117

 251 ACGGGCCCCTGAGCCCTACCCGCCTGGCCCCCACTATGGT.......... 288
     |||||||||           ||||||
 117 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 133

 289 CCAGAGGCGTGGGGTGAGCCG...CCTGCTGCGGCACCGGAAAGCCCAGC 335
     ::||||||||       |||      ||||||     |||||
 134 ProGlyAlaProGlyThrProGlyLeuProGlyLeuProGlySerProGl 150

 336 TCTTGCTGCTGTCAACCTGTCAACCTTTGGCCTGGAGTGTTTGCCGCA 385
     :::|||||||                      ||||||
 150 yAlaProGlyThrProGlyGlu.............GlyGlnG 160

 386 GGCATCACCTATGTGCCCCTGCTGCTGGAAGTGGGGGTAGAGGAGAA 435
     :::||||||      ||||||         |||
 160 lnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGly......... 173

 436 GTTCATGACCATGTGCGTGGGCATTGGTCCAGTGCCGGCCTGGTCTGTG 485
                       |||      ||||||||||:::|||
 174 ........................SerProGlyAlaProGly...Th 180

 486 TCCCCGGCTCTAGGCTCAGCCAGTGACCACTGGCGCTGGACGCTATGGCCGC 535
     ||:::||:::        :::
 180 rProGlyGln........................................ 184

 536 CGCCGGCCCTTCATCTGGGACACTGCCTTGGGCATCCTGCTGAGCCT... 582
     ||:::||||:::::|          ||||||
 185 ..........GlyLeuProGlySerProGlyAlaProGly 194

 583 CTTTCTCATCCCAAGGCCGGACCGGCGCCTGGCTAGCAGGGCTGCT.....GTGCC 626
     |||:::|||||                 |||:||||
 195 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211

 627 CGGATCCCAG.......GCCCCTGGAGCT..GGCACTGCTCATC 661
     ||| |||  ||||||||||       ||||||
 211 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr......P 226

 662 CTGGGCGTGGG.........GCTGCTGGACTTCGTGCTGGCCAGGTGTG 699
     ||||||:::|||                ||||
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/397,633A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 762 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-397-633A-26

alignment_scores:
Quality: 365.00   Length: 751
Ratio: 1.058   Gaps: 52
Percent Similarity: 45.939   Percent Identity: 30.093

alignment_block:
US-09-030-606-110 x US-08-397-633A-26

Align seg 1/1 to: US-08-397-633A-26 from: 1 to: 762

```
 84 GGTGCCCACAGCAGCAGGTGTG.....AGCATGGGCTGAGAAGCTGG 127
    |||::||| |||        ||| |||      ::||
 52 GlyThrProGlyProGlnGlyLeuProGlyGlySerProGlyAlaProGlyThr 68

128 ACCGGCACCAAAGGC.....TGGCAGAAATGGGCGCCTGGC........ 164
    |||::||:::||| |||      :::        ||||||||
 68 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG 85

165 ........TGATTCCTAGGCAGTTGCGGCAGCAAGGAGGAGAG 200
            |||||||| ||||:|| ::||::|||
 85 lyGluGlyGlnHisHisLeuGlyGly...AlaAlaGlnAlaGlyLysGlnAlaGlyAsp 100

201 GCCGCAGCTTCTGGACAGAGCGACGAGCGAAGCAGTTCTGGAGTGCCTGA 250
    :::::::|||      ||||||        :::||||:::|||
101 ValGlySerProGlyAlaProGlyThrProGlnGlyLeuProGl 117

251 ACGGCCCCCTGAGCCCTACCCGCCTGGCCCCACTATGGT.......... 288
    |||||||| |||              ::||||:||
117 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 133

289 CCAGAGGCTGTGGTGGGTGAGCCG...CCTGCTGCGGCACCGGAAAGCCCAGC 335
    |||  |||  |||  ||| :::    |||      ::||||||||:::
134 ProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGl 150

336 TCTTGTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGCCGCA 385
    ;||||||:::|||:::||:                    |||:|::
150 yAlaProGlyThrProGlyGlu.........................GlyGlnG 160

386 GGCATCACCTATGTGCCCGCCTCTGCTCGGAAGTGGGGTAGAGGAGAA 435
    ::||||||||| :::||||::  ||||||::   ::|||||:::
160 lnHisHisLeuGlyGlyAlaAlaLysGlnAlaGlyAspValGly....... 173

436 GTTCATGACCATGGTGCTGGGCATTGGTCCAGTGGTGCCCTGGCCTGCTGTG 485
    |||  :::||||||:::||  |||   ::||||:||||
174 ..............................SerProGlyAlaProGly..Th 180

486 TCCCGGCTCCTAGGCTCAGCAGCTCACCACTGCCGCTGGACGCGTATGGCCGC 535
    |||::||:::|||
180 rProGlyProGln.................................... 184
```

1509 TGTTCCT.................GCCCAAATACCGA 1528
     |:::||| ||| |||
 508 yThrProGlyGluGlyGlnHisHisLeuGlyGlyAla.LysGlnAla 524

1529 GGGGACACTGGAGGTGCTAGCAGTGAGGACGCCTGATGACCAGCTTCCT 1578
     ||||||| |||:::||::: || :||
 525 GlyAspValGlySerProGlyAlaProGlyThrProGlnGlyLe 541

1579 GCCAGGCCCT.....AAGCCTGGAGCTCCCTCCTAATGGACACGTGG 1622
     |||||| |||||||:::||| |||:::|||
 541 uProGlySerProGlyAlaProGlyThrProGlnGlyLeuProG 558

1623 GTGCTGGAGGCAGTGGCCTCTCCACCTCCACCCGCCTGCGGGGCC 1672
     ||::: |||:::: ||| ||| ::||| |||:::
 558 lySerProGlyAlaProGlyThrProGlnGlyLeuProGlySer 574

1672 1672
 575 ProGlyAlaProGlyThrProGlyGluGlyGlnHisHisLeuGlyGl 591

1673 ..TCTCCTGTGATGTCTCCGTACGTGGTGGTGAGCCCA 1715
   :::||| ||| ||||| |||||
 591 yAlaLysGlnAlaGlyAsp..............ValGly.SerPro 601

1716 CCGAGGCCAGGTGGTTCCGGGCCCGGGCATCTGCCTGGA.......... 1755
     |||||| |||::: |||||| ||||||||
 602 GlyAlaPro.GlyThrProGlnGlyLeuProGlySerProGlyA 618

1756 ..CCTCGCCATCCTGGA...TAGTGCCTTCCTGCTGTCCCAGGTGGCCCC 1800
      |||||:::||| |||||:::||::::||||||||:::|||
 618 laProGlyThrProGlnGlyLeuProGlySerProGlyAlaPro 634

1801 ATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCT 1850
     |||:::||| |||:::||
 635 GlyThrProGlyProGlnGlyLeuProGlySerPro.............. 646

1851 ATATGGTGTCTGCCCAGGCCTGGGGTCTGGTCGC......CATTTACTTT 1894
     |||::: |||:::|||:::|| ||||||||
 647 ...GlyAlaProGlyThrProGlyGluGlyGlnHisHisLeu..... 660

1895 GCTACACAGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTA 1944
     :::::: ||||||| |||
 661 ....GlyGly.....AlaAlaLysGlnAlaGlyAspValGlySerPro 672

1945 GAAAACTTCCAGCACATTGGGGTGGGTGAGGCCTGCCTCACTGGGTCCCAGC 1994
     ||| ||:::|||:::||||||||
 673 GlyAlaProGlyThrProGlnGlyLeuProGlySerProGlyAl 689

1995 TCCCGCTCCTGTTTAGCCCCATGGGGCTGCCGGGCTGGCCG 2035
     | |:::   :::|||:::|||||||||| |||
 689 a.ProGlyThrProGlnGlyLeuProGlySerPro 702

```
1336  ..TCGAGCAGTCTATTGGCCAGTGTGGCAGCTT....TCCCTGTGCTG 1379
         ::::::::::         |||:::         :::
 448  lyGlyAlaLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464

1380  CCGGTGCCACATGCCTGTCC....                 ACAGTGTG 1408
      |||     |||||||||||||                     |||||||
 465  ProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGl 481

1409  GCCGTGGTGACAGCTTCAGCGCCCTCACCGGGTCACCTTCTCAGCCCT 1458
      |||     |||  |||||||||||||::||:||: ::|||
 481  yPro.....GlnGlyLeuProGlySerProGlyAlaProGly.ThrPro 495

1459  GCAGATCCTGCCCTACACACTGGCTCCCTACCACCGGGAGAAGCAGG 1508
        |||||||||       |||||||    |||||     |||||
 496  GlyPro........GlnGlyLeuProGlySerProGlyAlaProGl 508

1509  TGTTCCT.....                        GCCAAATACCGA 1528
      |::|||
 508  yThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAla.LysGlnAla 524

1529  GGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCT 1578
      |||||||::||||::         :::::
 525  GlyAspValGlySerProGlyAlaProGlyThrProGlyGlnGlyLe 541

1579  GCCAGGCCCT.....AAGCTCGAGCTCCCTTCCTAATGGACACGTG 1622
      ||||||              ||||||||   |||:::|||    |
 541  uProGlySerProGlyAlaProGlyThrProGlyGlnGlyLeuProG 558

1623  GTGCTGGAGGCAGTGGCTGCTCCCACCTCCACCCGCCTGCGGGGCC 1672
      ||:::||   |||  |||||||||::::
 558  lySerProGlyProGlyThrProGlyGlnGlyLeuProGlySer 574

1672  .................                            1672
 575  ProGlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGl 591

1673  ......TCTGCCTGTGATGTCTCGCAGCGTGTGGTGGGTGAGCCCA 1715
            |::::|||       |:||        |||||||
 591                                   ValGly.SerPro 601

1716  CCGAGGCCAGGTGGTTCCGGGGCCGGCCATCTGCCTGA.......... 1755
      ||||||                                  |||||||
 602  GlyAlaPro.GlyThrProGlyProGlnGlyLeuProGlySerProGlyA 618

1756  ..CCTCGGCCATCCTGGA..TAGTGCTTCCTGCTGTCCCCAGGTGGCCCC 1800
        |||||      |||::|       ||||||||    ||||::|||:||
 618  laProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaPro 634

1801  ATCCCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCT 1850
                                            |||:|:::|||
 635  GlyThrProGlyProGlnGlyLeuProGlySerPro.............. 646

1851  ATATGGTGCTCGCCGCAGGCCTGGTCC......CATTTACTTT 1894
                                         ||||||
 647  ....GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeu.... 660

1895  GCTACACAGGTAGTATTGACAAGAGGACTTGGCCAAATACTCAGCGTA 1944
             |||:::||||||::||: |||||||||
 661  GlyGly.....AlaLysGlnAlaGlyAspValGlySerPro 672

1945  GAAAACTTCAGCACATTGGGGTGGAGGGCCTGCCTCCACTGGGTCCCAGC 1994
                    |||  ::||||||||       |||||||||
 673  GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 689

1995  TCCCCCGCCTCCTGTTAGCCCCATGGGGCTGCCGGGCTGGCCG 2035
      ||::::                 ||||||||||||::||||||| :||
 689  a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro 702
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-120

seq_documentation_block:
; Sequence 120, Application US/08642255
; Patent No. 5773249

```
 536  CGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGGCT.... 582
                      ::|     |||:::|||     |||
 185  .............GlyLeuProGlySerProGlyAlaProGly 194

 583  CTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTGCT.....GTGCC 626
      ||:::|||        :::|||     ::|||
 195  ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211

 627  CGGATCCCAG........GCCCCTGGAGCT...GGCACTGCTCATC 661
      |||                :::|||||||||  |||||
 211  oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr......P 226

 662  CTGGGCGTGGG.........GCTGCTGGACTTCTGTGGCCAGGTGTG 699
      |||||::|||                ::|||
 226  roGlyGlyGlyGlnGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAsp 242

 700  CTTCACTCCACTGGAGGGCCTGCTCTCTGACCCTTCCGGGACCC...... 744
      ::: |||  |||||||||||||||
 243  ValGlySerProGlyAlaProGly.........ThrProGlyProGlnGl 256

 745  .........GGACCACTGTCGCCAGGCCTACTCTGTCTAT 775
               :::|||   :::||||||
 256  yLeuProGlySerProGlyAlaProGlyThrProGly.......... 268

 776  GCCTTCATGATCAGTCTTGGGGGTGCCTGGGCTACCTCCTGCCTCCCTGCCAT 825
      |||||||||||                ::|||
 269  ..........ProGlnGlyLeuProGlySerProGlyAla.......... 278

 826  TGACTGGGACCACCAGTGCCCTGGCCCC.....CTACCTGGCACCCAGG 869
      ||||||                           |||||||
 278  ..........ProGlyThrProGlyGlnGlyLeuProGlySerPro.. 291

 870  AGGAGTGCCCTCTTTGGCCTGCTGCCTCAATCTTCCCTCACCTGCGTAGCA 919
      |||:::|||            ::: |||||
 292  .GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyA 308

 920  GCCACACTGCTGGTGGGTGAGGAGCCACCGAGCCAGC 969
      ||||||||||        ::|||||||||
 308  laLysGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324

 970  AGAAGGGCTGTCGGCCCCCTCCTTGTGCCCCACTGCGTGTCCATGCCGGG 1019
      ||:::|||        |||     ::: |||||
 325  ProGlnGlyLeuProGlySerProGlyAlaPro......GlyThrProGl 339

1020  CCCGCTTGGCTTTCCG...GAACCTGGGCGCCGGCCGCTTCCCCGGCTGCAC 1066
      |||:::|||        ::|||||||       |||:|||
 339  yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 355

1067  CAGCTGTGCTGCCGCCATGCCCCCCACCCTCCTTGTGCCCCACTGCGTGTCCATGCCGGCTGA 1116
      |||||                      |||||||
 355  roGlnGlyLeuProGlySerPro....GlyAlaProGlyThrProGlyGlu 370

1117  GCTGTGCAG.....CTGGATGGCACTCATGACCTTCACGCTGTT...... 1155
      ::: |||        ||| ::|||
 371  GlyGlnHisHisLeuGlyGlyAlaLysGlnAlaGlyAspValGlySe 387

1156  ......TTACACGGATTTCGTGGCGGAGGGGCTGTACCAGGGCGTGCCC 1198
            :::|||||||                 :::|||||||
 387  rProGlyAlaProGlyThrProGlyLeuProGlnGlyLeuProGlyProG 404

1199  AGAGGGCTGAGCCGGGCACCGGACCCGGAGCAACACTATGATGAAGGCGTTCG 1248
      ||:::|||                     |||||
 404  lyAla...ProGlyThrProGlyProGlnGlyLeu............... 414

1249  GATGGCCAGCCTGGGGCT.....GTTCCTGCAGTGCCGCCATCTCCCTGG 1292
      ::: |||||||              ||||||| |||||||
 415  ProGlySerProGlyAlaProGlyThrProGlyProGlyLeuProGl 431

1293  TCTTCTCTCTGGTCATGAGCCGGCCGGTGTGCAGCGATTCGGCAC....... 1335
      ||||||   |||  |||  :::|||       |||:||
 431  ySerProGlyAlaProGlyThrProGlyGlyGlnGlnHisHisLeuG 448
```

Wed Sep 29 14:26:53 1999

GENERAL INFORMATION:
```
APPLICANT: CAPPELLO, Joseph
APPLICANT: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 762 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
```
US-08-642-255-120

alignment_scores:
```
Quality:    364.00.          Length:      751
Ratio:       1.067           Gaps:         49
Percent Similarity: 45.406   Percent Identity: 29.694
```

alignment_block:
US-09-030-606-110 x US-08-642-255-120

Align seg 1/1 to: US-08-642-255-120 from: 1 to: 762

```
 84 GGTGCCCACCAGCAGCAGGTGTTG......AGCATGGGCTGAGAAGCTGG 127
    ||:::||  ||||||||||||       |||||||   :::
 52 GlyThrProGlnGlyProGlyLeuProGlySerProGlyAlaProGlyTh  68

128 ACCGGCACCAAAGGGC.......TGGCAGAAATGGGCGCCCTGCC....... 164
    |||::::||:||||||              :: |||||||||
 68 rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG  85

165 .............TGATTCCTAGGCAGTTGGCGCAGCAAGGAGGAGAG 200
               |||||||||::: ||| :::||||:::
 85 lyGluGlyGlnGlyHisHisLeuGlyGlyAlaArg...GlnAlaGlyAsp 100

201 GCCGCAGCTTCTGGAGCAGAGCCGAGACGAAGCAGTTCTGGAGTGCCTGA 250
    :::::::||  |||||          :: ::|||||::::||||
101 ValGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGl 117

251 ACGGCCCCCTGAGCCCTACCCGCCCTGGCCCACTATGGT.......... 288
    ||||||:::|| ||||||||           :: |||||
117 ySerProGlyAlaProGlyThrProGlyPro.GlnGlyLeuProGlySer 133

289 CCAGAGGCTGTGGGTGAGCCG..CCTGCTGCGGCACCGGAAAGCCCAGC 335
    |||  ||| |||  |||  :::  |||||| |||||||
134 ProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGl 150
```

```
336 TCTTGCTGGTCAACCTGCTAACCTTGGCCTGGAGTGTGTTGGCCCGCA 385
    |||  |||:::::
150 yAlaProGlyThrProGlyGlu................GlyGlnG 160

386 GGCATCACCTATGCGCCCCTCTGCTGCTGGAGTGGGGTAGAGGAGAA 435
    ::||||||||  ::::|||
160 lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGly......... 173

436 GTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTGGGCCTGGTCTGTG 485
    |||  ||||||  |||    :::::||||||
174 ...............SerProGlyAlaProGly...Th 180

486 TCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGC 535
    |||:::|||::
180 rProGlyProGln...................... 184

536 CGCCGGCCCTCATCTGGCACTGTCCTTGGCCATCCTGCTCTAGCCT... 582
    :::: ||| ||| :||||||
185 ...............GlyLeuProGlySerProGlyAlaProGly 194

583 CTTTCTCATCCCAAGGCCCGGCTGGCTAGCAGGGCTGCT.......GTGCC 626
    ||||||  ::|||    ::|||      :::||
195 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211

627 CGGATCCCAG...........GCCCCTGGAGCT...GGCACTGCTCATC 661
    ||||  |||             ::||||||||| |||||||
211 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.......P 226

662 CTGGGCGTGGG...........GCTGCTGGACTTCTGTGGCCAGGTGTG 699
    ||||:::||             ::||||||||
226 roGlyGluGlyGlnGlyHisHisLeuGlyGlyAlaArgGlnAlaGlyAsp 242

700 CTTCACTCCACTGGAGGCCCTGCTTCTGACCCTTCCGGGACCC........ 744
    ::: |||  :::||:||:    ||||||||
243 ValGlySerProGlyAlaProGly.........ThrProGlyProGln 256

745 ....................GGACCACTGTGCCAGGCCTACTCTGTCTAT 775
    :::|| ::||||||
256 yLeuProGlySerProGlyAlaProGlyThrProGly............ 268

776 GCCTTCATGATCAGTCTTGGGGCTGCTGGCTACCTTCCTGCCTGCCAT 825
    ||||||||||  ||||||||
269 .............ProGlnGlyLeuProGlySerProGlyAla...... 278

826 TGACTGGGACACCAGTGCCTGCCCCC.......CTACCTGGGCACCCAGG 869
    |||||||                   ||||||||||  |||
279 ..........ProGlyThrProGlyProGlnGlyLeuProGlySerPro. 291

870 AGGGATGCCTCTTGGCCTGCTCACCCTCATCTTCCTCACCTGCTAGCA 919
    ||:::||  |||    ::: :::||||     |||||
292 .GlyAlaProGlyThrProGlyGluGlyGlnGlyHisHisLeuGlyGlyA 308

920 GCCACACTGCTGCTGTGGCTGAGGAGGCAGCGCTGGGCCCCACCGAGCCAGC 969
    |||||  :::|||         :::::|||
308 laArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324

970 AGAAGGGCTGTCGCGCCCCCCTTGTCCCCCACTGCGTGTCCATGCCCGGG 1019
    ||||||||  |||     ||  :: ::||||||
325 ProGlnGlyLeuProGlySerProGlyAlaPro......GlyThrProGl 339

1020 CCCGCTTGCCTTCCG...GAACCTGGGCGCCCTGCTTCCCCGGCTGCAC 1066
     ||||::|||  ::|||    ::||||||       :::|||
339 yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 355

1067 CAGCTGTGCTGCGCAATGCCCCGCACCCTGGCCCGGGCTTCTTCGTGGCTGA 1116
     ||  |||||||    ::|||      |||||||
355 roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 370

1117 GCTGTGCAG..........CTGGATGGCACTCATGACCTTCACGGTGTT..... 1155
     ::: |||             :::||| |||||
371 GlyGlnGlyHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySe 387
```

```
1156 ......TTACACGGATTCGTGGCGAGGGGCTGTACCAGGGCGTGCCC 1198
              |||                  :::|||||||
387  rProGlyAlaProGlyThrProGlyInGlyLeuProGlySerProG  404

1199 AGAGCTGAGCCGGCCCGAGGCCCGGAGACAATGATGAAGGCGTTCG 1248
             |||      |||||||:::        |||
404  lyAla...ProGlyThrProGlyInGlyLeu..............  414

1249 GATGGGCAGCCTGGGGCT......GTTCCTGCAGTGCGCCATCTCCCTGG 1292
             ||||:||||||         ::|||:::  :::  |||||||
415  ProGlySerProGlyAlaProGlyThrProGlyInGlyLeuProGl  431

1293 TCTTCTCTGGTCATGACCGGCTGTGCAGCGATTCGGCACTCGAGCA 1342
               :::::|||                 ::|||
431  ySerProGlyAlaProGlyThrProGlyGluGlyInHisHisLeuG  448

1343 GTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGGCTGCCGGTGCCACATG 1392
              ::    ||:||:::                        |||
448  lyGlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr  464

1393 CCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCTCA......  1436
            |||  |||||    |||     ::  |||
465  ProGlyProGln...GlyLeuProGlySerProGlyAlaProGlyThrPr  480

1437 ..............CCCGGTCACCTTCTCAGCCCTGCAGATCCTGCC 1470
                         |||:::|||: ::|:|||:::  |||:::
480  oGlyProGlnGlyLeuProGlySerProGly.AlaProGlyThrProGly  496

1471 CTACACACTGCCCTCCCCTCCTACCACCGGAGAAGCAGGTGTTCCT...... 1515
             |||||||     ||||||
497  ...ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyGl  512

1516 ..............GCCCAAATACCGAGGGGACCACTGGA 1540
                         |||  ||| |||||||:::|||
512  uGlyGlnHisHisLeuGlyGlyAla.ArgGlnAlaGlyAspValGly  528

1541 GGTGCTAGCAGTGAGGACAGCCTGATGACCAAGCCTCCTGCCAGGCCCT.. 1588
            ::  :::     ||        ::: |||||||
529  SerProGlyAlaProGlyThrProGlyInGlyLeuProGlySerPr  545

1589 ....AAGCCTGGAGCTCCCTTCCCTAATGACACACGTGGCTGCTGGAGGCA 1634
           ||||:::||| |||||||         :::|||
545  oGlyAlaProGlyThrProGlyInGlyLeuProGlySerProGlyA  562

1635 GTGGGCCTGCTCCCCCACCTCACCCGCGCTCTGCGGGGCC............ 1672
            ||  |||     |||||||:::|||
562  laProGlyThrProGlyInGlyLeuProGlySerProGlyAlaPro  578

1673 ..............................TCTGC 1677
                                         :::|||
579  GlyThrProGlyGluGlyInGlnHisHisLeuGlyGlyAlaArgGlnAl  595

1678 CTGTGATGTCTCCGTACGTGTGTGGTGGTGAGCCCACCGAGGCCAGGG 1727
            |||||| ||||||| |||||||
595  aGlyAsp..............ValGly.SerProGlyAlaPro.Gl  605

1728 TGGTTCCGGGGCCGGGGCATCTGCCTGGA.............CCTCGCCATC 1765
            ||||| |||||||||                     |||
605  yThrProGlyInGlyLeuProGlySerProGlyAlaProGlyThrP  622

1766 CTGGA...TAGTGCCTTCCTGCTGTCCCAGGTGGCCCATCCCTGTTAT 1812
           ||||| |||||||:::|||
622  roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly  638

1813 GGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTG 1862
            ||:::|||                    ||:::|||
639  ProGlnGlyLeuProGlySerPro..............GlyAlaPr  649

1863 CCGCAGGCTGGGTCTCGGCCATTACTTTGCTACACAGTAGTATTT 1912
           :::||||:::||   ||||:::|||
649  oGlyThrProGlyGluGlyGln..............  656

1913 GACAAGAAGCGACTTGGCCCAAAATACTCAGCGTAGAAAACTTCCAGCACATT 1962
```

```
657  ..GlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerPro  672
                                            :::  |||||||
      |||!|::::||||||                     ::|

1963 GGGGTG...............GAGGGCCTGCCTCACTGGGTCCCAGC 1994
     |||:::                    :::|||||||||
673  GlyAlaProGlyThrProGlyInGlyLeuProGlySerProGlyAl  689

1995 TCCCCGCTCCGTTAGCCCCATGGGGCTGCCGGGCTGGCCG 2035
     | |||::: ::|||:::||||||||||   |||
689  a.ProGlyThrProGlyInGlyLeuProGlySerPro  702
```

seg_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-397-633A-31

```
seg_documentation_block
; Sequence 31, Application US/08397633A
Patent No. 5773577
GENERAL INFORMATION:
    APPLICANT: Cappello, Joseph
    TITLE OF INVENTION: PRODUCTS COMPRISING SUBSTRATES CAPABLE
    TITLE OF INVENTION: OF ENZYMATIC CROSS-LINKING
    NUMBER OF SEQUENCES: 105
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
        STREET: 4 Embarcadero Center, Suite 3400
        CITY: San Francisco
        STATE: California
        COUNTRY: USA
        ZIP: 94111-4187
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/397,633A
        FILING DATE:
        CLASSIFICATION: 530
    ATTORNEY/AGENT INFORMATION:
        NAME: Rowland, Bertram I
        REGISTRATION NUMBER: 20,015
        REFERENCE/DOCKET NUMBER: A-58848-1/BIR PROP-011-1
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 781-1989
        TELEFAX: (415) 398-3249
        TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 31:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 762 amino acids
        TYPE: amino acid
        STRANDEDNESS: single
        TOPOLOGY: linear
    MOLECULE TYPE: peptide
US-08-397-633A-31
```

```
alignment_scores:
    Quality:  364.00                    Length:  751
    Ratio:    1.067                      Gaps:    49
Percent Similarity: 45.406     Percent Identity: 29.694

alignment_block:
US-09-030-606-110 x US-08-397-633A-31

Align seg 1/1 to: US-08-397-633A-31   from: 1  to: 762
```

```
84   GGTGCCCCACAGCAGCAGGTGTTG.........AGCATGGGCTGAGAAGCTCG 127
     |||:::|||                        |||  |||     |||||||
52   GlyThrProGlyInGlyLeuProGlySerProGlyAlaProGlyTh  68

128  ACCGGACCAAAGGGC..........TGGCAGAAATGGGCGCCCTGGC....... 164
     |||:::::||::|||          :::   ||||||||
68   rProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProG  85
```

```
325  ProGlnGlyLeuProGlySerProGlyAlaPro......          ::  GlyThrProGl 339
1020 CCCGCTTGGCTTCCG...GAACCTGGGCGCCCCTGCTTCCCGGCTGCAC 1066
     |||| ||||| |||||          ::|||||  |||::||||   ||:|
339  yProGlnGlyLeuProGlySerProGlyAlaProGlyThrPro...GlyP 355
1067 CAGCTGTGCTGCCGCATGCCCCGCCACCCTGCGCGGCCTTCGTGGCTGA 1116
     |! !||||| |||    ::|||   ||||           ::: |||
355  roGlnGlyLeuProGlySerPro...GlyAlaProGlyThrProGlyGlu 370
1117 GCTGTGCAG......CTGGATGGCACTCATGAACCTTCACGCTGTT..... 1155
     ||| ||| |||       :::|||||::|||::||::|||
371  GlyGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAlaGlyAspValGlySe 387
1156 ........TTACACGGATTCGTGGGCAGGGGCTGTACCAGGGCGGTGCCC 1198
                 ||| ::||| :|||| |||| |||||| ||||:|||||||
387  rProGlyAlaProGlyThrProGlnGlyLeuProGlySerProGlySerProG 404
1199 AGAGCTGAGCCGGCACCGAGGCCGGAGACACTATGATGAAGGCGGTCG 1248
     ::: ||| |||||||::: ||
404  lyAla...ProGlyThrProGlnGlyLeu.............. 414
1249 GATGGGCAGCCTGGGGCT......GTTCCTGCAGTGCGCCATCTCCCTGG 1292
     |||::|||||||               ||||::||||| |||| ||||||
415  ProGlySerProGlyAlaProGlyThrProGlnGlyLeuProGl 431
1293 TCTTCTCTGGTCATGGACCCGGTGGTCAGCGATTCGGCACTCGAGCA 1342
     | ||| ||| ||| |::||||||  :||| |||||| :||||
431  ySerProGlyAlaProGlyThrProGlyGluGlyLnGlnHisHisLeuG 448
1343 GTCTATTGGCCAGTGTGGCAGCTTCCTGTGCTGCGGTGCCACATG 1392
     ::  |||||  ||||:|||::
448  lyGlyAlaArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThr 464
1393 CCTGTCCCACAGTGTGGCCGTGGTGGACAGCTTCAGCCGCCCTCA...... 1436
     |||  |||||| |||          ::  ||| |||
465  ProGlyProGln...GlyLeuProGlySerProGlyAlaProGlyThrPr 480
1437 ........CCGGGTTCACCTTCTCCAGCCCTGCAGATCCTGCC 1470
             ||||||||||||| |:::|||||: ||: :|| |::|
480  oGlyProGlnGlyLeuProGlySerProGly.AlaProGlyThrProGly 496
1471 CTACACACTGGCCTCCCCTCCTACCACCGGAGAAGCAGGTGTTCCT.. 1515
     ||||||   ::  ||||:||||||
497  ...ProGlyProGlySerProGlyLeuProGlnGlyThrProGlyGl 512
1516 ........GCCCAAATACCGAGGGGACACTGGA 1540
                     |||||||::::|||
512  uGlyGlnHisHisLeuGlyGlyAla..ArgGlnAlaGlyAspValGly 528
1541 GGTGCTAGCAGTGAGGACCAGCCTGATGACCAGCTTCCTGCCAGGCCCT.. 1588
     ::: ||||    ::  |||||||||||||
529  SerProGlyAlaProGlyThrProGlnGlyLeuProGlySerProPr 545
1589 ....AAGCTGGAGCTCCTCCTAATGGACACGTGGTGCTGGAGGCA 1634
             |||||::|||  :::||| |||
545  oGlyAlaProGlyThrProGlnGlyLeuProGlySerProGlyA 562
1635 GTGGGCCTGCTCCCACCTCCACCCGGCCTCTGCGGGGCC.......... 1672
     ::  ||| ||||||| ::|||| |||:||| |||||
562  laProGlyThrProGlyLeuProGlySerProGlyAlaPro 578
1673 ....................TCTGC 1677
568  GlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyAlaArgGlnAl 595
1678 CTGTGATGTCTCCTACGTGTGGTGGTGGGTGAGCCACCGAGGCCAGGG 1727
     |||        |||||| |||||||||| ::: |||||  ||| ||| ||
595  aGlyAsp.............ValGly.SerProGlyAlaPro.Gl 605
1728 TGGTTCCCGGGCCGGGCATCTGCCTGGA..........CCTCGCCCATC 1765
     |                                     |||
```

```
165  ..........TGATTCCTAGGCAGTTGGCGGCAGCAAGGAGGAGAG 200
     |||||||  ||| ||::||| :|
85   lyGluGlyGlnGlnHisHisLeuGlyGlyAlaArg...GlnAlaGlyAsp 100
201  GCCGCAGCTTCTGGACGCAGACGCCGAGACGAGTTCTGGAGTGCCTGA 250
     ::::::|||  |||||        :::  ||:::|||   ::||| |||
101  ValGlySerProGlyAlaProGlyThrProGlnGlyLeuProGlyLeuProGl 117
251  ACGGCCCCCTGAGCCCTACCCGCCTGGCCCACTATGGT.......... 288
     ||| ||||||||| |||||||
117  ySerProGlyAlaProGlyThrProGly.GlnGlyLeuProGlySer 133
289  CCAGAGGCTGTGGGTGAGCCG...CCTGCTGCGGCACCGGAAAGCCCAGC 335
     ||| ||| ||| ||     :::  |||  ||::|||::   ||||||
134  ProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGl 150
336  TCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTTGGCCGCA 385
     |||     |||   :::::|||||| |||||| |||||     :||||::
150  yAlaProGlyThrProGlyGlu..............GlyGlnG 160
386  GGCATCACCTATGTCCGCCTCTGCTGCTGGAAGTGGGGGTAGAGAGAA 435
     ::|||||||||       ||||||||    ||:||
160  lnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGly........ 173
436  GTTCATGACCATGGTCTGGGCATTGGTCCAGTGCTGCTGGGCCTGGTCTGTG 485
     |||   ||||||   ||:||||||| |||     ::::|||||
174  ..................SerProGlyAlaProGly...Th 180
486  TCCCGTCCTAGGCTCAGCCACCAGTGACCACTGCGCTGGACGCTATGGCCGC 535
     |||::||||::
180  rProGlyPro.GlnGlyLeuProGlySerProGlyAlaProGlyThr...P 226
536  CGCCGGCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCT... 582
     |||    ||    ||| :::|||||||
185  ..............GlyLeuProGlySerProGlyAlaProGly 194
583  CTTTCTCATCCCAAGGGCCGGCTGGTGCGGCTAGCAGGGCTGCT.....GTGCC 626
     ||:::|||  :::||| |||||||
195  ThrProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 211
627  CGGATCCCAG.........GCCCCTGGAGCT..GGCACTGCTCATC 661
     ||| |||  ::|||||||| ||||||
211  oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr......P 226
662  CTGGGCGTGGG.........GCTGCTGACTTCTGTGGCCAGGTGTG 699
     |||||||||       ||||      ||
226  roGlyGluGlyGlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAsp 242
700  CTTCACTCCACTGCGAGGGCCGTCCTCTGACCCTCTGACCTTCCGGGACCC.. 744
     |||||||    |||||||||:::
242  ValGlySerProGlyAlaProGly..........ThrProGlyProGlnGl 256
745  ...........GGACCACTGTCGCCAGGCCTACTCTGTCTAT 775
             ::|||| :|||||||
256  yLeuProGlySerProGlyThrProGly.............. 268
776  GCCTTCATGATCAGTCTTGGGGCTGCCTGGGCTACCTCCTGCCTGCCAT 825
     |||||:||| ||| ::||||||||||
269  ...........ProGlnGlyLeuProGlySerProGlyAla...... 278
826  TGACTGGGACACCAGTGCCCTGGCCCC.....CTACCTGGGCACCCAGG 869
     ||||||||| |||||||||
279  ..........ProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlySerPro.. 291
870  AGGAGTGCCTCTTTGCCCTTCACCCTCATCTTCCTCACCTGCCTAGCA 919
     |||::||| ::  ||||| |||||    ||::|||
292  .GlyAlaProGlyThrProGlyGluGlyGlnGlnHisHisLeuGlyGlyA 308
920  GCCACACTGCTGTGGCTGCGTGTGCGGAGGCAGCCTGGCCCCACCGAGCCAGC 969
     |:::: |||||| |||   ||: :|::|||
308  laArgGlnAlaGlyAspValGlySerProGlyAlaProGlyThrProGly 324
970  AGAAGGGCTGTCGGCCCCCTCCTTGTCGCCCCACTGCGTCCATGCCGCGGG 1019
```

```
                                    Ratio:        1.026      Gaps:              66
Percent Similarity:   40.506      Percent Identity:  30.495

alignment_block:
US-09-030-606-110/rev x US-08-642-255-62    .

Align seg 1/1  to: US-08-642-255-62  from: 1  to: 1064

2314 AAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     ::|||||||    |||||||::|||                 |||::::
  12 GluAsnPro.....GlyValThrGln.............. .....LeuAsnAr 21

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCGAGTCCCCATTCCAGTGCATGGAGCC 2215
     ::::::::|||||||        |||    :::||      :::|
  21 gLeuAlaAlaHisProProPheAlaSerAspProMetGlyAlaHisGlyP 38

2214 CTTCTGGCCTCCCTGTATAAGTCCAGACTGAAACCCCTGGAAGGCCTC 2165
     ||::|||                                 ::|||
  38 roAlaGly.............. ....... ....ProLysGlyAla 44

2164 CAG.............. ......TCAGGCAGCCCTAGACTGGGGAGAG...... 2135
     :::                      |||:::|||
  45 HisGlyProAlaGlyProLysGlyAlaProGlyProAlaGlyProProGl 61

2134 .AGGAGAGGGACGCCCCAGCCCCAGTGTGCAGCTACGCACCTCAGCAG 2086
      |||||  |||  |||||                            :::|
  61 ySerArgGlyAspProGlyProPro............. .........GlyA 71

2085 CACAGGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGG 2036
     ||  |||       |||  :::       ||||||::     :::
  71 laProGlyProAlaGlyProPro........ ...GlySerArgGlyAspPro 84

2035 CGGCCAGCCCGGCAGCCCCATGGGCTAACAGGAGCGGGAGCTGGGA.. 1988
     |||               |||  |||:::      |||||||
  85 GlyProProGlyAlaProGlyProAlaGlyProProGlySerArgGlyAs 101

1987 ............. ......CCCAGTGAGGCAGGCCCTCCACCCCCATGTG 1957
                         |||:::|   |||||||||||
 101 pProGlyProProGlyAlaProGlyProAlaGlyProPro.......... 114

1956 CTGGAAGTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTTGTCAAATAC 1907
     |||||||
 115 .Glyser.......... 116

1906 TACCTGTGTAGCAAAGTAAATGGCGACCAGAC.............. .....CCAGG 1870
                    ::|||||||       |||||
 117 ..............ArgGlyAspProGlyProProGlyAlaProGl 127

1869 CCTGCGGCAGACACCATATAGGCAGTGACAGACTGGCTGGACGTGAGCTGGACAAT 1820
         |||           ::: |||   |||
 127 yProAlaGlyProProGlySerArgGlyAspProGlyProProGlyAlaP 144

1819 GGAGCCCATAAACACAGGGATGGGGCCCACCTGGGACAGCAGGA...AGGCAC 1773
     :::|||                  ||||||||::: ||
 144 roGlyPro........ AlaGlyProProGlySerArgGlyAspProGly 157

1772 TATCCAGGATGGCGAGGTCCAGGCAGATGCCCCGGC.......... ....CC 1735
     |||||||       |||||::        |||||
 158 ProProGlyAlaProGlyProAlaGlyProGlySerArgGlyAspPr 174

1734 GGAACCACCCTGG...CCTCGGTGGCTCACCCACCACC.ACACGTACGG 1689
         |||||| |       |||    ||||||    :::|||      :
 174 oGlyProProGlyAlaProGlyProAlaGlyProSerArgGlyA 191

1688 AGACATCACAGGCAGAGAGGCCCCGCAGAGCCGGGTGGAGGTGGGAGCAGG 1639
         ::::|       |||||||                 |||||||||||
 191 spProGlyProProGlyAlaProGlyProAlaGlyProProGlySerArg 207

1638 CCACTGCCTCCAGCACCCACGTCCATTAGGAAGGAGCTCCAGGCTT 1589
         |||||||          |||      ||| |||||
 208 GlyAspProProGlyProProGlyAlaProGlyProProGlyProGlySe 224

1588 A......GGGCCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCCACTGCTAG 1545
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-62

```
 605 yThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrP 622

1766 CTGGA...TAGTGCCTTCCTGCTGTCCCAGGTGGCCCATCCCTGTTTAT 1812
     |||||       |||||||::: |||||||::|||
 622 roGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGly 638

1813 GGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTG 1862
              |||::::|||                 |||::::|||
 639 ProGlnGlyLeuProGlySerPro.............. .....GlyAlaPr 649

1863 CCGCAGCCTGGGTCTGGTCGGCCATTTACTTTGCTACACAGGTAGTATTT 1912
     |||:::|||:::|||                   :::|||     :::||
 649 oGlyThrProGlyGluGlyGln.............. .......... 656

1913 GACAAGAGCGACTTGGCCAAAATACTCAGCGTAGAAAACTTCCAGCACATT 1962
     |||:::||||||                  :::    :::
 657 ..GlnHisHisLeuGlyGlyAlaArgGlnAlaGlyAspValGlySerPro 672

1963 GGGGTG..........GGAGGCGTGCCTCACTGGGTCCCCAGC 1994
     |||:::                  ::|||||||||       ||
 673 GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAl 689

1995 TCCCCGGCTCCTGTTAGCCCCATGGGCTGCCGGGCTGGCCG 2035
     | |||::: :::||:::||||||||||| |||
 689 a.ProGlyThrProGlyProGlnGlyLeuProGlySerPro 702
```

alignment_scores:
;   Quality: 361.00   Length: 869

```
 224 rArgGlyAspPProGly...........................................  229
1544 CACCTCCAGTGTCCCCGTATTGGGCAGGAACACCTGCTTCTCCCGG  1495
 230 ..ProProGlyAlaPro.GlyPro...AlaGlyProProGlySerArgGl  244
1494 TGGTAGAGGAGGCCAGTGTGT......AGGGCAGGATCTGCAGGGCTGA  1451
 244 yAspPProGlyProProGlyAlaProGlyAlaGlyProProGlySerA  261
1450 GAAGGTGAACCCGGTGAGGGGCCGGCTGAAGCTGTCACCACGGC........  1409
 261 rg.GlyAspPProGlyProProGly.......AlaHisGlyProAlaGl  274
1408 .........CACACTGTGGACAGGCATGTGGCCACCGGCAGCCACA  1372
 274 yProLysGlyAlaHis.......GlyProAlaGlyProLysGlyAlaHis.  288
1371 GGG.........AAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT  1334
 289 GlyProAlaGlyProLysGlyAla.HisGlyPro...............  299
1333 GCCGAATCGCTGCACCAGCCGT...CCATGACCAGAGAGAGACCAGGG  1287
 300 .......AlaGlyProLysGlyAlaProGlyProProGly  313
1286 AGCCTAGGAGGCGGACAACAGACCAGGCGGCACTGGCAATGCCAGC  1237
 314 SerArgGlyAspPProGly..ProProGlyAlaPro............  324
1236 TCATAGTGTCTCGGGGCGTCCCGGCTCAGCTCTGGCACGCCCTG  1187
 325 ........GlyProAlaGlyProProGlySerArgGlyAspPProG  337
1186 GTACA......GCCCCTCGCCCACGAAATCCGTGTAAAACAGCGTGAAGG  1143
 337 lyProProGlyAlaProGlyProAlaGlyPro..............  347
1142 TCATGAGTGCCATCCAGCTGCGCACAGCTCAGCCAGCGAAGAGCCGGCCA.  1095
 348 ...........ProGlySer..ArgGlyAspPProGlyProProGlyAlaPr  360
1094 ...GGGTGCGGGGCATGCGGCAGCACAGCT............GGTG  1064
 360 oGlyAlaProGlyProGlySerArgGlyAspPProGlyProProGlyA  377
1063 CAGCCCGGGAAGCAGGGCGCGCCCAGGTTCC...........GGAAAGCCA  1026
 377 laProGlyProAlaProGlyProProGlySerArgGlyAspPProGlyProPro  393
1025 AGCGGGCCCGGCCATGGACACGACGGGGGACAAGG.......AGGGGGCC  982
 394 GlyAlaProGlyProAlaGlyProProGlySerArgGlyAspPProGlyPr  410
 981 GACAGCGCCCTTCGCTGCTGGGGGGCCCAGCGCTGCCT..........  942
 410 oProGlyAlaPro...........GlyProAlaGlyProGlySerA  423
 941 .........CCTCAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGA  900
 423 rgGlyAspPProGlyProProGly....................GlyAla  431
 899 TGAGGGTGAGCAGGCCAAAGAGGCACT..........CCTCCCTGG  865
 432 ProGlyProAlaGlyProProGlySerArgGlyAspPProGlyProProGl  448
 864 GTGCCCAGGTAGGGGG......CCAGGCACTGGGTGTCCCAGTCAATGGC  821
 448 yAlaProGlyProAlaGlyProProGly...................  457
 820 AGGCAGGAGGTAGCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGA  771
     :::.||||||||||| |||
```

```
 458 ..SerArgGlyAspPProGlyProPro..............................  465
 770 CAGAGTAGGCGCTGGCGACAGTGTCCGGTCCCGGTCCCGGAGAGAGGTCAG  726
 466 ...GlyAlaProGlyProAlaGlyProProGlySerArgGlyAspPProGl  481
 725 ......................AGAGCAGGGCGCCTCCAGTGAGTGAAGCACAC  695
 481 yProProGlyAlaProGlyAlaGlyProProGlySerArgGlyAspP  498
 694 CTGGCCACAGAAGTCAGCAGCCCACGCCCAGGATGAGCAGTGCCAGCT  645
 498 rOGly.....ProProGlyAlaProGly.......AlaGlyProPro  510
 644 CCAGGGGCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGGCCCTT  595
 511 GlySerArgGlyAspPProGlyProProGly......AlaHisGlyProAl  525
 594 GGGATGAGAA.......AGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCA  551
 525 aGlyProLysGlyAlaHisGlyProAlaGly...........ProL  537
 550 GATCGAAGGGCGCGGCGCCGCCATAGCGTCCACGCCAGTGGTCACTGGCTG  501
 537 ySGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly..........  550
 500 AGCCTAGGAGGCGGGACAACAGACAGGCCCAGCACTGGACCAATGCCCAGC  451
 551 ProAlaGlyProLysGlyAlaPro.............GlyPr  563
 450 ACCATGGTCATGAACTTCTCCCTCACCCCA...CTTCCAGCAGCAGAGG  404
 563 oProGlySerArgGlyAspPProGlyProProGlyAlaProGly.......  577
 403 CGGCACATAGGTGATGCCGCGGCCAAACACACCTCCAGGCCAAAGGTTA  354
 578 ....................ProAlaGlyProProGlySerArg...  585
 353 GCAGGTTGACCAGCAAGAGCTGGGCTTTCCGGTGCCCGCAGCA........  312
 586 ..GlyAspPProGlyProProGlyAlaProGlyProAlaGlyProProGl  601
 311 .............GGCGGCTCACCCAGCCCTCTGGACCATAGTGGGCCA  275
 601 ySerArgGlyAspPProGlyProProGlyAlaProGlyPro.AlaGlyPro  617
 274 GGGCGGGTAGGGCTCAGGGGGGCCCTTCAGCCACTCCAGAACTGCTTCGTCT  225
 618 Pro......GlySerArgGlyAspPProGlyProPro............  627
 224 CGGCTCGTCCAGAAGCTGCGGCCCTCGCCCTCCTGCTCCCGCCAACTGC  175
 628 .....GlyAlaProGlyProProGly.....................S  638
 174 CTAGGAATCAGCCAGGGCCCCATTTCTCGCACCCCCTTTGGTGCGCGGTCCA  125
 638 erArgGlyAspPProGly............ProProGlyAlaProGlyPro  650
 124 GCTTCTCAGCCCCATGCTCCA...............ACACCTGCTGCTGTGG  90
 651 AlaGly.ProProGlySerArgGlyAspPProGlyProProGlyAlaProG  667
  89 GGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTGCC.........  49
 667 lyPro...AlaGlyProGlySerArgGlyAspPProGlyProProGlyGly  682
  48 ...................CGAGGGCGCGGCCGTGTCACCCGGA  25
 683 AlaProGlyAlaGlyProProGlySerArgGlyAspPProGlyGly  697

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-642-255-132

seq_documentation_block:
```

```
; Sequence 132, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
;    APPLICANT: CAPPELLO, Joseph
;    APPLICANT: FERRARI, Franco A.
;    TITLE OF INVENTION: High Molecular Weight Collagen-Like
;    TITLE OF INVENTION: Protein Polymers
;    NUMBER OF SEQUENCES: 135
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET: 4 Embarcadero Center, Suite 3400
;       CITY: San Francisco
;       STATE: California
;       COUNTRY: USA
;       ZIP: 94111-4187
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/642,255
;       FILING DATE:
;       CLASSIFICATION: 435
;    ATTORNEY/AGENT INFORMATION:
;       NAME: ROWLAND, Bertram I.
;       REGISTRATION NUMBER: 20,015
;       REFERENCE/DOCKET NUMBER: A55556-3/BIR
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (415) 494-8700
;       TELEFAX: (415) 494-8771
;       TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 132:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH: 829 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-642-255-132
```

```
alignment_scores:
           Quality: 357.00          Length:     829
             Ratio: 0.994             Gaps:      59
Percent Similarity: 43.305  Percent Identity: 30.881

alignment_block:
US-09-030-606-110/rev x US-08-642-255-132

Align seq 1/1  to: US-08-642-255-132  from: 1  to: 829
```

```
2320 GCTCCCAAAAACCCTTCTCTAGGTGTGTCT...CAACTAGGAGGCTAGCT 2274
     |||||  ::|||:||:|||  ||  :::   :||||||||||
  32 AlaProGlyThrProGlyGluGlyGlyGlnHisHisLeuGlyGlyAlaLy 48

2273 GTTAACCCTGAGCCTGGGTAATCCACCTGAGAGTCCCCGCCATTCCAGTG 2224
     ::  :::::||||:::|||  ||  :::
  48 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThrPro....... 62

2223 CATGGAGCCCCTTCTGGCCCTCCCCTGTATAAGTCCAGACTGAAACCCCTTG 2174
     :::|||||:::||||||||||||
  63 ....GlyProGlnGlyLeuPro................... 68

2173 GAAGGCCTCCAGTGAGCCTCAGCAGCCCTAGAGACTGGGGAGGAGGAGGAC 2124
     |||||||||                          |||::
  69 ...........GlySerPro......................GlyAl 73

2123 GGCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCACGCACAGGGTGGCA 2074
     ||| |||::  :::||||||| :::  ||||||:::
  73 aProGlyThrProGlyGln...........GlyLeuProGlySerP 86

2073 GCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGG 2024
     |||
```

```
  86 roGlyAlaProGlyThrProGly..ProGlnGlyLeuProGlySerProG 102
     |||              |||        |||           ::|||||||
2023 CAGCCCCATGGGCGCTAACAGAGGCGGGAGCTGGGACCCAGTGAGGCAGG 1974

 102 lyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAla 118
     :::|||||          :::|||                 |||
1973 CCCTCCACCCCATGTGCTGGAAGTTTCTACGCTGAGTATTTGGCCAAG 1924

 119 ProGlyThrPro..................................... 122
     |||::|||||
1923 TCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGCGACCAGACC 1874

 123 ....GlyProGlnGlyLeuPro...................GlySerP 131
     :::  ||| ||||||                          :::||
1873 CAGGCCTGCGGCAGACACCATATAGGCAGTGACAGACGACTGGCTGAGCTGGA 1824

 131 roGlyAlaProGlyThrProGlyProGlnGlyLeuProGly......... 144
     |||||  |||||||                  |||
1823 CAATGGAGCCCATAAACAGGGATGGGCCACCTGGACAGCAGGAAGGCA 1774

 145 ......SerPro..........GlyAlaProGlyThrProGlyProGl 155
     |||||              ||  |||||||||     |||   ::
1773 C.................TATCCAGGATGGCGAGGTCCAGGCAGAT 1745

 155 nGlyLeuProGlySerProGlyAlaProGlyThrProGlnGlyL 172
     :                                          |||||
1744 GCCCCGGC..............CCGGAACCACCCTGGCCTCGGTG....... 1714

 172 euProGlySerProGlyAlaProGlyThrProGlnGlyLeuPro 188
     |||||              |||     |||:::
1713 GGCTCACCCACCACCACGTACGGAGACATCACAGGCA......GAGGC 1670

 189 GlySerProGlyAlaProGlyThrProGlnGlyLeuProGlySe 205
     |||||||||  :::            |||     ||||||     :::
1669 CCCGCAGAGGCGCGGTGAGGTGGAGCAGGCCACTGCCTCCAGCACCCA 1620

 205 rProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProg 222
     |||              |||        |||           :::|||
1619 CGTGTCCATTAGGAGGGAGGAGCCCCAGGTTAGGCCTGCAGGAAGAAGCTG ::: 1570

 222 lyAlaProGlyThrProGlyProGlnGlyLeu..ProGly.SerProGl 237
     ::::||              |||        |||     |||  |||
1569 GTCATCAGGCTGTCCTCCACTGCTAGCACCTCCAGTGTCCCCTCGGTATT. 1521

 237 yAlaProGlyThrProGly.......................ProGlnGlyLeuP 248
     :::  |||  |||                          |||  |||::::
1520 ....TGGCAGGAACACCTGCTTCTCCCGGTGTGAGAGGCGGAGGCCAGTG 1476

 248 roGlySerProGlyAlaProGlyThrProGlnGlyLeuPro... 263
     |||::::||:||||||||     :::|||||  ::::
1475 TGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGTGAGGGCGGCT 1426

 264 .........GlySerProGlyAla.ProGlyThrProGlnGlyL 277
     |||||              |||           |||
1425 GAAGCTGTCACCACGGCCACACTGGGACAGGCATGTCGCACC...GGC 1379

 277 euProGlySerProGly..............AlaProGlyThrProGly 288
     |||              |||||        |||     |||
1378 AGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGA 1329

 289 ..ProGlnGly...LeuProGlySerPro.................. 296
     |||           |||  |||
1328 ATCGGCTGCACCAGCCGGTCCATGACCAGGAGAGAGACCAGGGAGATGGCCG 1279

 297 ...GlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGl 312
     :::|||||||:::  ::: |||     |||||
1278 CACTGCAGGAACA..........................GCCC 1262

 312 yAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaP 329
     ::: |||              |||        |||           :::|||
1261 CAGGCTGCCCATCC...GAACGCCTTCATCATAGTGTCTCCGGGCCTCGG 1215
```

```
374 ACACCTCCAGGCCAAAGGTTAGCAGGTGACCAGCAAGAGCTGGCTTTC 325
    ||||||       |||    |||    |||::::   ||
586 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 602

324 CGGT...GCCGCAGCAGGCGGCTCACCCACAGCCTCTGGACCATAGTGGG 278
    :||          :::              |||      |||      |||
602 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.ProGly 618

277 CCAGGCGGG...TAGGGCTAGCAGGGGCCGTTCAGGCCACTCCAGAACTGCT 231
    |||   |||   |||          |||||| |||    |||||
619 ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl 635

230 TCGTCTCGGCTCGCTCCAGAAGCTGCGGCCGCCTCTCCCTTGCCGCCGCC 181
    |||   :::::|||   |||   |||         :::::|||
635 nGlyLeuProGlySerProGlyAlaProGlyThrProGly.......... 648

180 AACTGCCTAGGAATCAGCCAGGCGCCCATTCTGCCAGCCCCTTTGGTGCC 131
    |||::::||   |||||||||      :: ||
649 .....ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrPro 662

130 GGTCCAGCTTCTCAGCCCCATGCTCA............ACACCTGCTG 96
    ||||::    ||| :::                       |||||||||
663 GlyProGlnGly.LeuProGlySerProGlyAlaProGlyThrProGlyP 679

95 CTGTGGGGGCACCCAGTGGGACACGTCATCACTCAGATCCTGGC... 49
    |||   |||    |||   :::   :::     |||||
679 roGlnGly...LeuProGlySerProGlyAlaProGlyThrProGlyPro 694

48 CGAGGCGCCGGCGTGTCACCCGGAGCC 22
    ||||||||||||||||
695 GlnGlyLeuProGlySerProGlyAla 703
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-397-633A-53

```
seq_documentation_block:
; Sequence 53, Application US/08397633A
; Patent No. 5773577
; GENERAL INFORMATION:
;   APPLICANT: Cappello, Joseph
;   TITLE OF INVENTION:  PRODUCTS COMPRISING SUBSTRATESCAPABLE
;   TITLE OF INVENTION:  OF ENZYMATIC CROSS-LINKING
;   NUMBER OF SEQUENCES: 105
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;     STREET:  4 Embarcadero Center, Suite 3400
;     CITY:  San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94111-4187
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/397,633A
;     FILING DATE:
;     CLASSIFICATION:  530
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Rowland, Bertram I
;     REGISTRATION NUMBER:  20,015
;     REFERENCE/DOCKET NUMBER:  A-58848-1/BIR PROP-011-1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 781-1989
;     TELEFAX:  (415) 398-3249
;     TELEX:  910 277299
; INFORMATION FOR SEQ ID NO: 53:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 829 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
```

```
329 roGlyThrProGlnGlyLeuProGlyAlaProGly 345
    |||   |||   ||| ||||||||
1214 TGCCGGCTCAGCTCTGGGCACGCCCTGGTACAGCCCCTCGCCCACGAAA 1165

346 ThrProGlyPro...GlnGlyLeuProGlySerProGlyAlaProGlyTh 361
    |||   |||:::   ||||||||
1164 TCCGTGTAAAACAGGCTGAAGGTCATGAGTGCCATCCAGC.......... 1125

361 rProGlyPro............LeuProGlyProGlnGly 374
    :||   ::::||
1124 ..TGCCAGCTCAGCCACGAAGAGCCCGGCCAGGTGCGGGGCATGCGGC 1077

374 roGlyThrProGlnGlyLeuProGlySerProGlyAlaProGly 390
    |||   |||   ||| :::::  ||||||
1076 AGCCAGCT......GGTGCAGCCCGGGAGCAGGCAGGCGCCCAGGTTCCGG 1033

391 ThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThrPr 407
    :||::|||   ||||||||
1032 AAAGCCAAGGGGG...CCCGGCATGGACACAGCAGTGGGGGACAAGGAGGG 986

407 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProG 423
    |||:::||   ||||||    :::|||
985 GGCCGACAGCCCTTCTGCGGCTCGGTGGGCCCAGCCGTCGCT...CCT 939

423 lyProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyPro 439
    |||| :::||   ::||  ||||||:::||  ||||||||
938 CAGCCACCACAGTGTGGCTGCTACGGCAGGTGAGGAAGATGAGGGTGAGC 889

440 Gln................................GlyLeuProGlySerPr 446
    |||                                  ||||||
888 AGGCCAAAGAGGCACTCCT.............CCTGGGTGCCCAGGTA 854

446 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGly. 462
    |||||||||||||
853 GGGGGCCAGGGCACTCGGTGTCCCAGTCAATGCAGGCAGGAGGTAGCCCA 804

463 ..AlaProGly.......ThrProGlyProGlnGlyLeuProGlySerPro 476
    ::||||||      |||||||||
803 GGCAGCCCCCAAGACTGATCATGAGGCATAGACAGTAGGCCTGGCGA 754

477 GlyAlaPro.................................GlyThrProGlyPr 484
    ::: |||||                                 |||
753 CAGTGT...CCGGGT...CCCGGAGAGGTCAGAGAGCAGGGCCTCCAG 710

484 oGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnG 501
    :::|||   |||||| ||||||  ||||||:::||
709 TGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGCCCAGGA 660

501 lyLeuProGlySerProGly.....AlaProGlyThrProGlyProGln 515
    ||||||:::::||||||      ::::::|||  ||||||
659 TGAGCAGTGCCAGCTCCAGGGGCCTGGATCCGGGCACAGCAGCCCTGCT 610

516 GlyLeuProGlySerProGly.....ProGlyThrProGly..... 528
    ||| ||||||||||||||
609 AGCCAGCCGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGA 560

529 .........ProGlnGly...LeuProGlySerProGlyAlaProGlyT 541
    |||   :::|||  ||||||:::::|| ||||||:::||
559 CAGTG...CCCAGATGAAGGGCCGGCCATTAGCGTCCACGCCAGT 513

541 hrProGlyProGlnGlyLeuProGlyGly 549
    ||   ||:::|||
549

512 GGTCACTGGCTGAGCCTAGGAGCCGGACACAGAGACCACTGGA 463
    |||::|| ::||||||||
550 ..........SerProGlyAlaProGlyThrProGlyProGln..... 560

609 AGCCAGCCGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGA 560
    ||   ||| ||||||:::::|| ||||||:::||
529 .........ProGlnGly...LeuProGlySerProGlyAlaProGlyT 541

462 CCAATGCCCACCACCATGGTCATGAACTTCTCCTCTACCCCCA...... 420
    :::||||||
561 ......GlyLeuProGlySerProGlyAlaProGlyThrProGlyProG 575

419 .....CTTCCAGCAGCAGAGGCGGCACATAGGTAGTGCCTGCGGCCCAAAC 375
    ||||||:::::|| ::::||
575 lnGlyLeuProGlySerProGlyAla......ProGly 585
```

us-09-030-606-110.rai

Wed Sep 29 14:26:53 1999

```
;  MOLECULE TYPE:  protein
US-08-397-633A-53

alignment_scores:
           Quality:   357.00          Length:   829
             Ratio:    0.994            Gaps:    59
Percent Similarity:    43.305  Percent Identity:  30.881

alignment_block:
US-09-030-606-110/rev x US-08-397-633A-53

Align seg 1/1  to: US-08-397-633A-53  from: 1  to: 829

2320 GCTCCCAAAAACCCTTCTCTAGGTGTGTCT..CAACTAGGAGGCTAGCT 2274
        ::||||  .::||||.  ::||||||||||
  32 AlaProGlyThrProGlyGluGlyLnGlnHisLeuGlyGlyAlaLy 48

2273 GTTAACCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATTCCAGTG 2224
        :::   :::||||||:  |||  :: |||
  48 sGlnAlaGlyAspValGlySerProGlyAlaProGlyThrPro........ 62

2223 CATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAGACTGAAACCCCCCTTG 2174
        :::||||||  ||||||
  63 ....GlyProGlnGlyLeuPro...................... 68

2173 GAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGGAGGAGGAC 2124
        |||||||||
  69 ...........................GlySerPro..........GlyAl 73

2123 GCCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCAGCAGGGTGGCA 2074
        :||  |||::||  |||
  73 aProGlyThrProGlyPro...................GlyLeuProGlySerP 86

2073 GCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGGGCGCCAGCCCCG 2024
        ||  |||  ||  ::||||||
  86 roGlyAlaProGlyThrProGly..ProGlnGlyLeuProGlySerProG 102

2023 CAGCCCCATGGGACTAACAGAGAGCGGGGAGCTGGGACCCAGTGAGGCAGG 1974
        ::|||||||        ::| |||
 102 lyAlaProGlyThrProGlnGlyLeuProGlySerProGlyAla 118

1973 CCCTCACCCCCAATGTGCTGGAAGTTTTCACGCTGAGTATTGGCCAAG 1924
        |||.::|||||
 119 ProGlyThrPro................................... 122

1923 TCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACC 1874
        |||  ||| ||||||                        ::|
 123 ....GlyProGlnGlyLeuPro................GlySerP 131

1873 CAGGCCTCGCGCAGACACCATATAGGCAGTGACAGCTGGCTGAGCTGGA 1824
        ||||||  |||
 131 roGlyAlaProGlyThrProGlyLeuProGly.......... 144

1823 CAATGGAGCCCATAAAACAGGGATGGGGCCACCTGGGACAGCAGGAAGGCA 1774
        ::|||||||  |||  ::||||
 145 .....SerPro........GlyAlaProGlyThrProGlyProGl 155

1773 C...................TATCCAGGATGGCGAGGTCCAGGCAGAT 1745
        |||||||||  .: |||
 155 nGlyLeuProGlySerProGlyAlaProGlyThrProGlnGlyLeuL 172

1744 GCCCCCGGC.........CCGGAACCACCCTGGCCTCGGTG........ 1714
        ||||||  .:: ||||::
 172 euProGlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuPro 188

1713 GGCTCACCCACCACCACCTACGGAGACATCACAGGCA......GAGGC 1670
        ||||||||                           ::
 189 GlySerProGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySe 205

1669 CCCGCAGAGCGCGGTGGAGTGGGAGCAGGCCACTGCCTTCCAGCACCCA 1620
        :||  |||  ::: ||||||  ::||||
 205 rProGlyAlaProGlyThrProGlnGlyLeuProGlySerProG 222
```

```
1619 CGTGTCCATTAGGAGGAAGGAGGTCCAGGCTTAGGGCTGGCAGGAAGCTG 1570
        ::::||                  ||||||  :::
 222 lyAlaProGlyThrProGlnGlyLeu...ProGly.SerProGl 237

1569 GTCATCAGGCTGTGTCTCACTGCTAGCACCTCCAGTGTCCCTCGGTATT 1521
        :::  |||  |||                 ||||| ||||::
 237 yAlaProGlyThrProGly...............ProGlnGlyLeuP 248

1520 ....TGGGCAGAAACACCTGCTTCTCCCGGTGGTGAAGAGGGAGGCCAGTG 1476
        :::||||||         ::: |||  |  |||||
 248 roGlySerProGlyAlaProGlyThrProGlnGlyLeuPro... 263

1475 TGTAGGGCAGGATCGTCAGGGCTGAGAAGGTGAACCCGTGAGGGCGGCT 1426
        |||||| |||  ::| |||::
 264 ........GlySerProGlyAla.ProGlyThrProGlnGlyLyL 277

1425 GAAGCTGTCACCACGGCCACACTGTGGGACAGGCATGTGGCACC...GGC 1379
        |||  ||||| |||
 277 euProGlySerProGly.................AlaProGlyThrProGly 288

1378 AGCCAGGGAGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGA 1329
        |||||||||  |||
 289 ..ProGlnGly...LeuProGlySerPro................... 296

1328 ATCGGCTGCACCAGCCGGTCCATGACCAGAGAAGACCAGGAGGATGGCG 1279
        ::||||||::: ||:::  ||::|
 297 ...GlyAlaProGlyThrProGlnGlyLeuProGlySerProGlyGl 312

1278 CACTGCAGGAACA...................................GCCC 1262
        ||  ||||||
 312 yAlaProGlyThrProGlyGlyGlyPro...............GlyAlaP 329

1261 CAGGCTGCCATCC...GAACGCCTTCATCATAGTGTCTCCGGGCCTCGG 1215
        ::: |||||||                     :::::|||||
 329 roGlyThrProGlyProGlnGlyLeuProGlySerProGlyAlaProGly 345

1214 TGCCCGGCTCAGCTCTGGGCACGCCCTGGTACAGCCCCTCGCCCACGAAA 1165
        :||:  |||||||    |||  |||  |||
 346 ThrProGlyPro...GlnGlyLeuProGlySerProGlyAlaProGlyTh 361

1164 TCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGC......... 1125
        ::::||:::||||
 361 rProGlyProGlnGly................LeuProGlySerProGlyAlaP 374

1124 ..TGCACAGCTCAGCCACGAAGACCGCGCAGGGTCGCGGGCATGGGGC 1077
        |||  ||||::                     ::||| |||
 374 roGlyThrProGlnGlyLeuProGlySerProGlyAlaProGly 390

1076 AGCACAGCT...........GGTGCAGCGGGGAAGCAGGCGGCCAGGTTCCGG 1033
        ::: ::                |||  |||  ||||||
 391 ThrProGlyAlaProGlyThrProGlySerProGlyAlaProGlyThrPr 407

1032 AAAGGCCAAGCGGG...CCCGGCATGGACAGCAGCTGGGGCGGACAAGGAGGG 986
        ||||:||      ::||
 407 oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr...ProG 423

985 GGCCCACAGCCCTCTGCTGGCTCGGTGGGGCCCAGCCGTGCCT...CCT 939
        ||||||      ::||| |||  ||||||
 423 lyProGlySerProGlyAlaProGlyThrProGlyPro 439

938 CAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAATGAGGGTGAGC 889
        :::|||||                   ||||||
 440 Gln.................................GlyLeuProGlySerPr 446

888 AGGCCAAAGAGGCACTCCT...............CCTGGGTGCCAGGTA 854
        ||||||||            ::||
 446 oGlyAlaProGlyThrProGlyProGlnGlyLeuProGlySerProGly. 462

853 GGGGGCCAGGGCCGGGTGGGAGTGGGAGCAGGCCAGGAGGTAGCCCA 804
        :::||||||                    ||||||||
 463 ..AlaProGly........ThrProGlyProGlnGlyLeuProGlySerPro 476
```

```
803  GGCAGCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCGA  754
     |||:||:|||                        :::  |||||||
477  GlyAlaPro.......                         ...GlyThrProGlyPr  484

753  CAGTGGT...CCGGGT...CCCGGAAGAGGTCAGAGAGCAGGGCCCTCCAG  710
     ::::||  |||||| ||||||                    ||||||
484  oGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGlnG  501

709  TGGAGTGAAGCACACCTGGCCACAGAAGTCAGCAGCCCCACGCCCCAGGA  660
     ::::||:|||||||      :::||||:||||||||       |||
501  lyLeuProGlySerProGly.....AlaProGlyThrProGlyProGln  515

659  TGAGCAGTGCCAGCTCCAGGGCCTGGGATCCGGCCACAGCAGCCCCTGCT  610
     :::|||||||||||        ||||:||||  :::
516  GlyLeuProGlySerProGlyAla......ProGlyThrProGly.....  528

609  AGCCAGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGA  560
          |||  |||||   |||||||  |||:::|||   |||||
529  ........ProGlnGly...LeuProGlySerProGlyAlaProGlyT  541

559  CAGTG...CCCAGATGAAGGGCCGGCGGGGCGGCCATAGCCGTCCACGCCAGT  513
     ||                                     |||
541  hrProGlyProGlnGlyLeuProGly.....                    549

512  GGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCCCAGCACTGGA  463
         |||  ||||||                     ||||||||
550  .........SerProGlyAlaProGlyThrProGlyProGln.....    560

462  CCAATGCCAGCACCATGGTCATGAACTTCTCTCTACCCCCA.......    420
                         |||  :::|||       |||
561  .......GlyLeuProGlySerProGlyAlaProGlyThrProGlyProG  575

419  .....CTTCAGCAGCAGAGGCGGCACATAGGTGATGCCTGCGGCCAAAC  375
     |||||||:::||||||                    ::::||
575  lnGlyLeuProGlySerProGlyAla...........ProGly  585

374  ACACCTCCAGGCCAAAGGTTAGCAGGTTGACCGACCAAGAGCTGGGCTTTC  325
     ||||||     |||  ||| |||::: ||:||
586  ThrProGlyGlnGly.LeuProGlySerProGlyAlaProGlyThrPr   602

324  CGGT...GCCCGAGCAGGCGGCTCACCCACAGCCTCTGGACCCATAGTGGG  278
     |||         ||||||||||||  |||  |||           |||
602  oGlyProGlnGlyLeuProGlySerProGlyAlaProGlyThr.ProGly  618

277  CCAGGCGGG...TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAAACTGCT  231
     |||   |||       ||||||  |||:::||||||||
619  ProGlnGlyLeuProGlySerProGlyAlaProGlyThrProGlyProGl  635

230  TCGTCTCGGCTCTGCTCCAGAAGCTGCGGCCTCTCCCTTGCTGCCGCC    181
     |||  |||||||| |||   :::|||:||
635  nGlyLeuProGlySerProGlyAlaProGlyThrProGly.........  648

180  AACTGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTGGTGCC  131
     |||:::||: |||||||||   :::  |||
649  ....ProGlnGlyLeuProGlySerProGlyAla...ProGlyThrPro  662

130  GGTCCAGCTTCTCAGCCCATGCTCA...............ACACCTGCTG  96
     |||||| |||  |||                         |||||||||
663  GlyProGlnGly...LeuProGlySerProGlyAlaProGlyThrProGlyP  679

95   CTGTGGGGGCACCTCAGTGGGGACACGTCTCATCACTCAGATCCTGGC...  49
     ||||||||||||||  :::  |||    |||   ||||||
679  roGlnGly...LeuProGlySerProGlyAlaProGlyThrProGlyPro  694

48   CGAGGGCGCCGGCTGTCACCCGGAGCC  22
     :::||| ||||||||||||
695  GlnGlyLeuProGlySerProGlyAla  703
```

This Page Blank (uspto)

```
                Copyright   GenCore version 4.5
                Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      September 28, 1999, 10:58:40 ; Search time 2928.69 Seconds
                                 (without alignments)
                                 3702.971 Million cell updates/sec

Title:       US-09-030-606-110
Perfect score: 3410
Sequence:    1 GGGAACCAGCCTGCACGCGC.........AAAAAAATAAAAAAAAAAAA 3410

Scoring table: IDENTITY_NUC

Searched:    679419 seqs, 1590154680 residues

Database :   GenEmbl:*
             1:  gb_bal:*
             2:  gb_ba2:*
             3:  gb_om:*
             4:  gb_ov:*
             5:  gb_pat:*
             6:  gb_ph:*
             7:  gb_pl1:*
             8:  gb_pl2:*
             9:  gb_prl:*
             10: gb_pr2:*
             11: gb_pr3:*
             12: gb_ro:*
             13: gb_st:*
             14: gb_sts:*
             15: gb_sy:*
             16: gb_un:*
             17: gb_vi:*
             18: em_fun:*
             19: em_htg:*
             20: em_hum1:*
             21: em_hum2:*
             22: em_in:*
             23: em_om:*
             24: em_or:*
             25: em_ov:*
             26: em_pat:*
             27: em_ph:*
             28: em_pl:*
             29: em_ro:*
             30: em_sts:*
             31: em_sy:*
             32: em_un:*
             33: em_vi:*
             34: gb_htg1:*
             35: gb_htg2:*
             36: gb_in1:*
             37: gb_in2:*
             38: em_ba1:*
             39: em_ba2:*
             40: em_hum3:*
             41: em_hum4:*
             42: gb_pr4:*
```

SUMMARIES

No.					ID	Description
2	101.2	3.0	1252	10	HSSTA	X82234 H.sapiens m
3	95.8	2.8	1544	4	OLANNEXM3	Y11254 O.latipes m
4	94.6	2.8	688	5	I03321	I03321 Sequence 12
5	93.4	2.7	1694	10	HSM800167	AL049382 Homo sapi
6	93.2	2.7	1794	5	E02349	E02349 cDNA sequen
7	92.8	2.7	625	8	AF049925	AF049925 Petunia x
8	92.8	2.7	2556	42	AF125948	AF125948 Homo sapi
9	92.2	2.7	1500	10	HSM800108	AL050024 Homo sapi
10	91.8	2.7	1500	4	XLNAKATP	Y11587 X.laevis mR
11	91.6	2.7	446	37	AF146743	AF146743 Mesobuthu
12	91.4	2.7	2638	7	ATAJ0058	AJ000058 Arabidops
13	91.2	2.7	1887	10	HSGLCNACT	Z82022 H.sapiens m
14	91	2.7	1887	42	AF118274	AF118274 Homo sapi
c 15	90.8	2.7	13414	5	A58523	A58523 Sequence 27
c 16	90.8	2.7	10288	5	A58524	A58524 Sequence 28
17	90.8	2.7	6671	5	AR011880	AR011880 Sequence 1
18	90.8	2.7	6671	5	I50973	I50973 Sequence 1
19	90.8	2.7	6671	5	I69486	I69486 Sequence 1
20	90.8	2.7	1483	10	HSM800068	AL049283 Homo sapi
c 21	90.8	2.7	8414	15	CVU30496	U30496 Cloning vec
c 22	90.8	2.7	8392	15	CVU30497	U30497 Cloning vec
23	90.6	2.7	3905	5	A03736	A03736 H.sapiens m
24	90.6	2.7	3905	5	A31790	A31790 H.sapiens m
25	90.4	2.7	837	10	S78214	S78214 APC=tumor s
26	90.2	2.6	1412	37	AF083228	AF083228 Caenorhab
27	90	2.6	1368	10	HSY16645	Y16645 Homo sapien
28	89.8	2.6	1474	5	A65341	A65341 Sequence 64
29	89.6	2.6	2462	10	HSM800419	AL050116 Homo sapi
30	89.6	2.6	705	10	HSM800237	AL049452 Homo sapi
31	89.6	2.6	1448	12	AF067728	AF067728 Rattus no
32	89.6	2.6	3581	12	RNSTOP	X93495 R.norvegicu
33	89.4	2.6	4670	42	AF104032	AF104032 Homo sapi
34	89.4	2.6	1559	4	OLJ000937	AJ000937 Oryzias l
35	89.4	2.6	144	5	I89947	I89947 Sequence 26
36	89.4	2.6	5959	10	HSHBRM	X72889 H.sapiens h
37	89.4	2.6	580	10	HSM800144	AL050277 Homo sapi
38	89.4	2.6	3482	10	HSM800550	AL050393 Homo sapi
39	89.4	2.6	3747	10	HSU42766	U42766 Human neuro
40	89.4	2.6	4915	12	HS2PHSAL2	X98834 H.sapiens m
41	89.4	2.6	1591	12	AF087943	AF087943 Rattus no
42	89.2	2.6	2134	4	XLCCTG	X84990 X.laevis Cc
43	89.2	2.6	868	8	AF049923	AF049923 Petunia x
44	89.2	2.6	1008	8	AF117707	AF117707 Lycopersi
45	89.2	2.6	1429	11	AF097996	AF097996 Homo sapi

ALIGNMENTS

```
RESULT  1
HSPE54C06/c
LOCUS       HSPE54C06      406 bp      DNA      STS      18-NOV-1998
DEFINITION  H.sapiens flow-sorted chromosome 1 HindIII fragment, SC1pE54C06,
            sequence tagged site.
ACCESSION   AL033941
NID         g3893713
VERSION     AL033941.1      GI:3893713
KEYWORDS    STS; Anonymous marker; single read.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
AUTHORS     Gregory,S., Kettleborough,R., Langford,C., Ross,M.T. and Hunt,S.E.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-1998) The Sanger Centre, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk
COMMENT     Vector: pBS1ISK+
            Marker stSG33426FS (Primer A : TGTAGGGCAGGATCTGCAG; Primer B :
            CTCGAGCAGTCTATTTGGCC; amplimer size : 141 bp) was mapped to
            chromosome 1 using Radiation Hybrid
            panel Genebridge 4 (GB4).
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	261.4	7.7	406	14	HSPE54C06	AL033941 H.sapiens

```
FEATURES        Location/Qualifiers
     source     1..406
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="1"
                /sex="Female"
                /dev_stage="adult"
                /tissue_type="EBV lymphoblastoid cell line"
                /clone_lib="SClpE"
                /clone="SClpE54C06"
BASE COUNT      97 a    128 c    115 g     66 t
ORIGIN

Query Match       7.7%;   Score 261.4;  DB 14;  Length 406;
Best Local Similarity 96.1%;  Pred. No. 5.4e-31;
Matches 268;  Conservative  0;  Mismatches 11;  Indels  0;  Gaps  0;

Qy 1237 TGAAGGCGGTTCGGATGGGCAGCGTCGGGCGTGTTCCTGCAGTGCGCCATCTCCCTGGTCTT 1296
        || ||||||||||||||||||||||||||||| ||||||||||||||||||||| |||||||
Db  395 TGCAGGCGGTTCGGATGGGCAGCGTCGGGCGTCTTCCTGCAGTGCGCCATCTCCGTGGTCTT 336

Qy 1297 CTCTCGGTCATGGACCGGCTGGTCGCAGCGATTCGGCACGTCGCAGCGAGTCTCATTGGCCAG 1356
        || ||||||||||||||||||||||||||||||||||||||  ||| |||| || |||||||
Db  335 CTATGTGGTCATGGACCGGCTGGTCGCAGCGATTCGGCACTTCGGAGCAGTTTATTTGGCCAG 276

Qy 1357 TGTGGCAGCTTTCCCTGTGGCGTGCCCGTGCCACATCCGTGTCCCACAGTGTGGCCGTGGT 1416
        |||||||||| ||||||||||| || || ||||||| | ||||||||||||||||||||||
Db  275 TGTGGCAGCATTCCCTGTGGCCGTGGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGT 216

Qy 1417 GACAGCTTCAGCCGCCCTCACCACGGGTTCACCTTCTCAGCCCCTGCAGATCCTGCCCTACAC 1476
        ||||||||||||||||||||||||||||||| || ||| |||||||||||||||||||||||
Db  215 GACAGCTTCAGCCGCCCTCACCACGGGTTCACGTTTCAGCCCCTGCAGATCCTGCCCTACAC 156

Qy 1477 ACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCT 1515
        |||||||||||| |||||||||||||||||||||| || |
Db  155 ACTGGCCTCCCTATACCACCGGGAGAAGCAGGTAGTCAT 117

RESULT 2
HSSTA
LOCUS       HSSTA       1252 bp    mRNA          PRI     05-JAN-1995
DEFINITION  H.sapiens mRNA for emerin.
ACCESSION   X82434
NID         g600618
VERSION     X82434.1   GI:600618
KEYWORDS    emerin; STA gene.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1252)
  AUTHORS   Bione,S., Maestrini,E., Rivella,S., Mancini,M., Regis,S., Romeo,G.
            and Toniolo,D.
  TITLE     Identification of a novel X-linked gene responsible for
            Emery-Dreifuss muscular dystrophy
  JOURNAL   Nat. Genet. 8 (4), 323-327 (1994)
  MEDLINE   95201795
REFERENCE   2  (bases 1 to 1252)
  AUTHORS   Toniolo,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-1994) D. Toniolo, Istituto di Genetica Biochemica
            ed Evoluzionistica, CNR, Via Abbiategrasso 207, 27100 Pavia, ITALY
FEATURES             Location/Qualifiers
     source          1..1252
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /cell_type="teratocarcinoma"
                     /cell_line="Human teratocarcinoma"
                     /map="Xq28"
     gene            59..823
                     /gene="STA"
```

```
     CDS             59..823
                     /gene="STA"
                     /codon_start=1
                     /product="emerin"
                     /protein_id="CAA57817.1"
                     /db_xref="PID:g600619"
                     /db_xref="GI:600619"
                     /db_xref="SWISS-PROT:P50402"
                     /translation="MDNYADLSDTELTTLLRRYNIPHGPVVGSTRRLYEKKIFEYETQ
                     RRRLSPPSSAASYSFSDLNSIRGDADMYDLPKKEDALLYQSKGYNDDYYEESYFTI
                     RTYGEPESAGPSRAVRQSVTSFPDADAFHHQVHDDDLLSSSEECKDRERPMYGRDSA
                     YQSITHYRPVSASRSSLDLSYYPTSSSTSFMSSSSSSWLTRRAIRPENRRAPGAGLG
                     QDRQVPLWGQLLFLVFVVLFPIYHFMQAEEGNPF"
BASE COUNT      313 a    344 c    312 g    283 t
ORIGIN
     polyA_signal    1125..1131

Query Match          3.0%;   Score 101.2;  DB 10;  Length 1252;
Best Local Similarity 87.9%;  Pred. No. 1e-06;
Matches 109;  Conservative  1;  Mismatches 14;  Indels  0;  Gaps  0;

Qy 3287 TGTTTATGTGACAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAA 3346
        ||||| ||| |||   |||  ||| || ||||  |||||||||||||||||||||||
Db 1111 TTTTTGTGGACACACAATAAAGCCCGTTATTTGTTAAAAAAAAAAAAAAAAAAAAAAAA 1170

Qy 3347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAA 3406
        ||||||||||||||||||:||||||||||||||||||||||  ||||||
Db 1171 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1230

Qy 3407 AAAA 3410
        ||||
Db 1231 AAAA 1234

RESULT 3
OLANNEXM3
LOCUS       OLANNEXM3      1544 bp    mRNA                    VRT    11-MAY-1999
DEFINITION  O.latipes mRNA for annexin max3.
ACCESSION   Y11254
NID         g3288569
VERSION     Y11254.1   GI:3288569
KEYWORDS    annexin max3; calcium-binding protein; phospholipid-binding
            protein.
SOURCE      Japanese medaka.
ORGANISM    Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
            Atherinomorpha; Cyprinodontiformes; Adrianichthyoidei;
            Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 1544)
  AUTHORS   Osterloh,D., Wittbrodt,J. and Gerke,V.
  TITLE     Characterization and developmentally regulated expression of four
            annexins in the killifish medaka
  JOURNAL   DNA Cell Biol. 17 (10), 835-847 (1998)
  MEDLINE   99025617
REFERENCE   2  (bases 1 to 1544)
  AUTHORS   Osterloh,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-FEB-1997) D. Osterloh, University Of Muenster, Dept.
            Of Medical Biochemistry, Von-Esmarch-Strasse 56, D-48149 Muenster,
            FRG
FEATURES             Location/Qualifiers
     source          1..1544
                     /organism="Oryzias latipes"
                     /db_xref="taxon:8090"
                     /dev_stage="embryo"
                     /clone="medaka annexin max 3"
     CDS             60..1073
                     /function="calcium- and phospholipid-binding protein"
                     /note="cDNA contains flanking EcoRI and XhoI sites"
                     /codon_start=1
                     /product="annexin max3"
```

```
                    /protein_id="CAA72124.1"
                    /db_xref="PID:e311341"
                    /db_xref="PID:g3288570"
                    /db_xref="GI:3288570"
                    /db_xref="SPTREMBL:O93346"
                    /translation="MSFIQAFLQQTVYLGMPDDSVLRNEGTVTAAPNFSPSGDAAVLD
                    KAIKAKGVDENTIIEILVKRSNEQRQQIKEAYQQASGKPLESALKSALKGDLEEVVLA
                    LLKTPAQYDAQQLKLAMKGLGTDEDTLIEILASRNNRQIMDLKKAYKEDYKKDLEEDI
                    RSDTSGDFRAALLALCKAGRTEGISEQLIDSDARALYEAGEGRKGKDCSVFIEILTTR
                    SGPHLRKVFERYSKYSKVDMAKAIDLEMKGDIESCLTAIVKCSGSRAFFAEKLNLAM
                    KGKGTRKNLTRIMVARSEIDMKLIKEEYKKNYGKTLYKDILDDTKGDYEKILLALCG
                    GDN"
BASE COUNT     547 a    308 c    354 g    335 t
ORIGIN

Query Match                      2.8%;  Score 95.8; DB 4;  Length 1544;
Best Local Similarity   91.7%;   Pred. No. 6.7e-06;
Matches  100;   Conservative     1; Mismatches     8;  Indels     0;  Gaps     0;

Qy    3302  AATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3361
            ||  |||||| |||| ||| || | ||||||||||||||||||||||||||||||||||
Db    1409  AAATAAAGCTTTTCAAATGATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1468

Qy    3362  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAA 3410
            ||||||| ::||||||||||||||||||||||||||||||||||||||||
Db    1469  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1517

RESULT 4
I03321
LOCUS       I03321        688 bp ss-DNA              PAT       05-MAR-1993
DEFINITION  Sequence 12 from Patent US 4886747.
ACCESSION   I03321
NID         g270715
VERSION     I03321.1  GI:270715
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Derynck,R.M.A. and Goeddel,D.V.
  TITLE     Nucleic acid encoding TGF-.beta. and its uses
  JOURNAL   Patent: US 4886747-A 12 12-DEC-1989;
            Genentech, Inc.;
            South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="unknown"
BASE COUNT     240 a    137 c    138 g    173 t
ORIGIN

Query Match                      2.8%;  Score 94.6; DB 5;  Length 688;
Best Local Similarity   90.1%;   Pred. No. 1.1e-05;
Matches  100;   Conservative     1; Mismatches    10;  Indels     0;  Gaps     0;

Qy    3300  AAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3359
            || ||||| ||||||| ||||||||| |||||||||||||||||||||||||||||||
Db     571  AACAATAAAGGCCCTATTCTAAGGTGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 630

Qy    3360  AAAAAAARAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAA 3410
            ||||||| :::||||||||||||||||||||| ||||||||||
Db     631  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 681

RESULT 5
HSM800167
LOCUS       HSM800167     1694 bp mRNA              mRNA        21-MAY-1999
DEFINITION  Homo sapiens mRNA; cDNA DKFZp586K1318 (from clone DKFZp586K1318).
ACCESSION   AL049382
NID         g4500170
VERSION     AL049382.1  GI:4500170
```

```
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1694)
  AUTHORS   Wambutt.R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct-Submission
  JOURNAL   Submitted (21-MAY-1999) MIPS, Am Klopferspitz 18a D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); sequenced by AGOWA within the cDNA
            sequencing consortium of the German Genome Project. This clone is
            available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..1694
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586K1318"
                     /clone_lib="586 (synonym: hute1). Vector pSport1; host
                     DH10B; sites NotI + SalI/MluI"
                     /dev_stage="adult"
                     /tissue_type="uterus"
     polyA_signal    1578..1583
     polyA_site      1599
BASE COUNT     672 a    238 c    218 g    566 t
ORIGIN

Query Match                      2.7%;  Score 93.4; DB 10; Length 1694;
Best Local Similarity   82.2%;   Pred. No. 1.5e-05;
Matches  106;   Conservative     1; Mismatches    22;  Indels     0;  Gaps     0;

Qy    3282  TATATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAATGTTTAAAAAAAAAAAA 3341
            |||| || ||||||| | ||| ||| ||| ||||  ||||||||||||||||||||||||||
Db    1558  TATATTGTTTACAATTAAAAATAAATTTACTACTTAATCCTAAAAAAAAAAAAAAAAAAA 1617

Qy    3342  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAA 3401
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1618  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1677

Qy    3402  AAAAAAAA 3410
            ||||||||
Db    1678  AAAAAAAA 1686

RESULT 6
E02349
LOCUS       E02349        1794 bp RNA               RNA
DEFINITION  cDNA sequence coding for eta peptide.
ACCESSION   E02349
NID         g2170584
VERSION     E02349.1  GI:2170584
KEYWORDS    JP 1990111796-A/1.
SOURCE      Bos taurus.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Artiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 1794)
  AUTHORS   Isobe,T., Okuyama.N. and Takahashi,Y.
  TITLE     PROTEIN ACTIVATING AMINO ACID HYDROXYLASE AND GENE THEREOF
  JOURNAL   Patent: JP 1990111796-A 1 24-APR-1990;
            TONEN CORP
COMMENT     OS   (Bovine)
            PN   JP 1990111796-A/1
            PD   24-APR-1990
            PF   21-OCT-1988 JP 198264097
            PI   ISOBE TOSHIAKI, OKUYAMA NORIO, TAKAHASHI YASUO PC
            C07K15/06,C07K13/00,C07K15/12,C12N15/12,C12P21/02//A61K37/24, PC
            C07K3/02;
            CC   strandedness: Single;
```

```
CC        topology: Linear;
CC        hypothetical: No;
CC        anti-sense: No;
CC        *source: tissue_type=cerebellum;
FH   Key             Location/Qualifiers
FT   5'UTR           1..165
FT   CDS             166..906
FT                   /product='Eta peptide'
FT   3'UTR           907..>1794
FT   polyA_signal    1671..1676.
FEATURES             Location/Qualifiers
     source          1..1794
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
BASE COUNT      506 a    454 c    471 g    363 t
ORIGIN

Query Match              2.7%;  Score 93.2;  DB 5;  Length 1794;
Best Local Similarity   69.4%;  Pred. No. 1.6e-05;
Matches 125;  Conservative    1;  Mismatches   54;  Indels    0;  Gaps    0;

Qy   3231 CTTATTATTTAGCGCGGGTGAATAATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTT 3290
          |||||| || |||  |||||||  ||| ||      |  || ||| |||| |   |||
Db   1601 CATGTCAATCAAGATGGGTGATTATGAAATGCCAGAATGCCTAAATTAAATGTTTTGGAAT 1660

Qy   3291 TATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAA 3350
          | |  |||  ||| ||| ||| ||   |  || |||||||||||||||||||||||||||
Db   1661 TCAGTGGGTAAAATAAAATGCTGCTTTGGGGATATTAAAAAAAAAAAAAAAAAAAAAAAAA 1720

Qy   3351 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
          ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1721 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1780

RESULT  7
AF049925
LOCUS       AF049925       625 bp   mRNA            PLN        02-JAN-1999
DEFINITION  Petunia x hybrida PGP214 (PGP214) mRNA, partial cds.
ACCESSION   AF049925
NID         g4105787
VERSION     AF049925.1  GI:4105787
KEYWORDS    .
SOURCE      Petunia x hybrida.
  ORGANISM  Petunia x hybrida
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Guyon,V., Astwood,J.D. and Taylor,L.P.
  TITLE     Isolation and characterization of petunia germinating pollen cDNAs
            expressed in conditionally male fertile pollen
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 625)
  AUTHORS   Guyon,V., Astwood,J.D. and Taylor,L.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1998) Genetics and Cell Biology, Washington State
            University, Science Hall, Pullman, WA 99164-4234, USA
FEATURES             Location/Qualifiers
     source          1..625
                     /organism="Petunia x hybrida"
                     /db_xref="taxon:4102"
                     /tissue_type="germinating petunia pollen treated with
                     kaempferol"
     gene            1..625
                     /gene="PGP214"
     CDS             <1..313
                     /gene="PGP214"
                     /note="A, R, L rich protein"
                     /codon_start=2
                     /product="PGP214"
                     /protein_id="AAD02551.1"
```

```
                     /db_xref="PID:g4105788"
                     /db_xref="GI:4105788"
                     /translation="DWIEHTKARHERAKEQGEEQNGLAPQEARRANHRAQLLARLAAR
                     RGAIVCIFRRDLPWITRYEARGASYSAPRANFDVIFRLGLGLTIYQAFVNPINTLLLS
                     F"
BASE COUNT      216 a    112 c    134 g    163 t
ORIGIN

Query Match              2.7%;  Score 92.8;  DB 8;  Length 625;
Best Local Similarity   91.5%;  Pred. No. 2e-05;
Matches  97;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy   3305 TAAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3364
          ||| ||| ||  |||  |||  |||||||||||||||||||||||||||
Db    519 TAATGGGTTACTTATTCTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 578

Qy   3365 AAAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
          |||:|||||||||||||||||||||||||||| |||||||||||||||
Db    579 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 624

RESULT  8
AF125948
LOCUS       AF125948      2556 bp    mRNA            PRI        12-APR-1999
DEFINITION  Homo sapiens DNA repair protein RAD52 delta isoform (RAD52) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF125948
NID         g4581005
VERSION     AF125948.1  GI:4581005
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2556)
  AUTHORS   Kito,K., Wada,H., Yeh,E.T.H. and Kamitani,T.
  TITLE     Identification of Novel Isoforms of Human RAD52
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2556)
  AUTHORS   Kito,K., Wada,H., Yeh,E.T.H. and Kamitani,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-1999) Internal Medicine, University of Texas,
            6431 Fannin, MSB4200, Houston, TX 77030, USA
FEATURES             Location/Qualifiers
     source          1..2556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1..2556
                     /gene="RAD52"
     CDS             44..400
                     /gene="RAD52"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA repair protein RAD52 delta isoform"
                     /protein_id="AAD24575.1"
                     /db_xref="PID:g4581006"
                     /db_xref="GI:4581006"
                     /translation="MSGTEEAILGGRDSHPAAGGSVLCFGQCYTAEEYQAIQKALR
                     QRLGPEYISSRMAGGQQKVCYIEGHRVINLANEMFGYNGWAHSITQQNVGEYALQQWG
                     LLHCPAPAESLLWVRR"
BASE COUNT      749 a    581 c    638 g    588 t
ORIGIN

Query Match              2.7%;  Score 92.8;  DB 42;  Length 2556;
Best Local Similarity   76.7%;  Pred. No. 1.8e-05;
Matches 112;  Conservative    1;  Mismatches   33;  Indels    0;  Gaps    0;

Qy   3265 TAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTT 3324
          |||| ||| ||||| |  |||||| |||| |  |||| |||| |||  ||| ||| |||
Db   2403 TTATTAACAAAATAGAAGAATATATTGCTTATCTGAAATGCTTGAAATGCTTGAAAAAAAA 2462
```

```
Qy    3325  TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAA  3384
            ||||||||||||||||||||||||||||||||||||||||||| ::||||||||||||||||
Db    2463  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2522

Qy    3385  AAAAAAAAAAAAAAAATAAAAAAAAAAAAA  3410
            ||||||||||||||||| ||||||||||||
Db    2523  AAAAAAAAAAAAAAAAAAAAAAAAAAA  2548
```

RESULT 9
HSM800108
LOCUS   HSM800108   1500 bp   mRNA   PRI   21-MAY-1999
DEFINITION Homo sapiens mRNA; cDNA DKFZp564D206 (from clone DKFZp564D206).
ACCESSION   AL050024
NID   g4884093
VERSION   AL050024.1 GI:4884093
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1500)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-MAY-1999) MIPS, Am Klopferspitz 18a D-82152
            Martinsried, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); sequenced by AGOWA within the cDNA
            sequencing consortium of the German Genome Project. This clone is
            available at the RZPD in Berlin.
            Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES   Location/Qualifiers
     source   1..1500
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp564D206"
                     /clone_lib="564 (synonym: hfbr2). Vector pAMP1; host
                     Xl-2blue; sites NotI + SalI"
                     /dev_stage="fetal"
                     /tissue_type="brain"
     gene   1..405
                     /gene="DKFZp564D206"
     CDS   <1..405
                     /gene="DKFZp564D206"
                     /note="unknown"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB43243.1"
                     /db_xref="PID:e1454921"
                     /db_xref="PID:g4884094"
                     /db_xref="GI:4884094"
                     /translation="GGCRELRSTKYISDGQCTSISPLKELVCAGECLLPVLPNWIGG
                     GYGTKYWSRRSSQEWRCVNDKTRTORIQLQCCDGSTRTYKITVVTACKCKRYTRQHNE
                     SSHNFEMSPAKPVQHRERKRRASKSSKHSMS"

                                Query Match   2.7%;   Score 92.2; DB 10; Length 1500;
                                Best Local Similarity 83.7%;   Pred. No. 2.4e-05;
                                Matches 103;   Conservative 1;   Mismatches 19;   Indels 0;   Gaps 0;

Qy   3288 GTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAA 3347
            |||||||||||||||||||| ||| ||| | ||||||||||||||||
Db   1375 GTTTCTTTAACAAAAATAAACACTGCTAAAAGTTAAAAAAAAAAAAAAAAAAAAAAAA 1434

Qy   3348 AAAAAAAAAAAAAAAAAAAÀRAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAA 3407
            ||||||||||||||||||| ::|||||||||||||||||||||| |||||||||
Db   1435 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1494
```

polyA_signal   1390..1395
polyA_site     1409

BASE COUNT    516 a    288 c    269 g    427 t
ORIGIN

```
Qy    3408  AAA  3410
            |||
Db    1495  AAA  1497
```

RESULT 10
XLNAKATP
LOCUS   XLNAKATP   1500 bp   mRNA   VRT   15-AUG-1997
DEFINITION X.laevis mRNA for Na,K-ATPase gamma subunit.
ACCESSION   Y11587
NID   g2052283
VERSION   Y11587.1   GI:2052283
KEYWORDS   ATPase; gamma subunit; gamma subunit of sodium potassium ATPase;
            NA+, K+-ATPase.
SOURCE   African clawed frog.
  ORGANISM Xenopus laevis
            Eukaryota; Metazoa; Chordata; Vertebrata; Amphibia; Batrachia;
            Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 1500)
  AUTHORS   Beguin,P., Wang.X., Firsov,D., Puoti,A., Claeys,D.,
            Horisberger,J.D. and Geering,K.
  TITLE   The gamma subunit is a specific component of the Na,K-ATPase and
            modulates its transport function
  JOURNAL   EMBO J. 16 (14), 4250-4260 (1997)
  MEDLINE   97392454
REFERENCE   2 (bases 1 to 1500)
  AUTHORS   Beguin,P.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-FEB-1997) P. Beguin, University of Lausanne,
            Institute of Pharmacology and Toxicology, Rue du Bugnon 27, 1005
            Lausanne, Switzerland
FEATURES   Location/Qualifiers
     source   1..1500
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /cell_line="A6"
                     /tissue_type="kidney"
     CDS   63..248
                     /codon_start=1
                     /product="gamma subunit of Na,K-ATPase"
                     /protein_id="CAA72326.1"
                     /db_xref="PID:e307198"
                     /db_xref="PID:g2052282"
                     /db_xref="GI:2052284"
                     /db_xref="SWISS-PROT:O13001"
                     /translation="MADAQDDMSQMQDKFTYDYETIRKGGLIFAAIAFVVGMLIIFSG
                     RFRCGRKKQLRALNDDM"

BASE COUNT   559 a   249 c   260 g   432 t
ORIGIN

                                Query Match   2.7%;   Score 91.8; DB 4;   Length 1500;
                                Best Local Similarity 81.4%;   Pred. No. 2.7e-05;
                                Matches 105;   Conservative 1;   Mismatches 23;   Indels 0;   Gaps 0;
```

```
Qy    3282  TATAATGTTTATGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAA  3341
            |||| || ||| |||  ||||||||||||||| ||  ||  |     ||||||||||||||
Db    1312  TATACTGTATATTGAATTATTGAATTATTAAAGGAGGCTGTGTTCTCAAAAAAAAAAAAAA  1371

Qy    3342  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAA  3401
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    1372  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1431

Qy    3402  AAAAAAAA  3410
            ||||||||
Db    1432  AAAAAAAA  1440
```

RESULT 11
AF146743
LOCUS   AF146743   446 bp   mRNA   INV   17-JUN-1999
DEFINITION Mesobuthus martensii BmK2 mRNA, complete cds.

```
ACCESSION   AF146743
NID         g5081730
VERSION     AF146743.1   GI:5081730
KEYWORDS    .
SOURCE      Manchurian scorpion.
ORGANISM    Mesobuthus martensii
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
            Buthoidea; Buthidae; Mesobuthus.
REFERENCE   1  (bases 1 to 446)
AUTHORS     Wenxin,L. and Shunyi,Z.
TITLE       Direct Submission
JOURNAL     Submitted (29-APR-1999) Virology Dept., Wuhan University, Luojia
            Mountain, Wuhan, Hubei Province, People's Republic of China
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Mesobuthus martensii"
                     /db_xref="taxon:34649"
                     /tissue_type="venom gland"
     CDS             7..180
                     /codon_start=1
                     /product="BmK2"
                     /protein_id="AAD39511.1"
                     /db_xref="PID:g5081731"
                     /db_xref="GI:5081731"
                     /translation="MKPRVFFLLFLLVAAMIETGESEENEEGSNESGKSTEAKNTDAS
                     VDNEDSDIDGDSD"
BASE COUNT    260 a     35 c     59 g     92 t
ORIGIN

Query Match           2.7%;   Score 91.6; DB 37;   Length 446;
Best Local Similarity  80.3%;  Pred. NO. 3.2e-05;
Matches 106;  Conservative   1;  Mismatches  25;  Indels   0;  Gaps   0;

Qy  3279  GAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAA  3338
          ||||||||||||||||||||||||||| | |||||| ||| ||||||||||||||||||
Db   239  GAAAAAATCTTTTACTGAAATGAAGTGAATAAATAATTATTTGCATAAAAAAAAAAAAA   298

Qy  3339  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT  3398
          ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db   299  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   358

Qy  3399  AAAAAAAAAAA  3410
          |||||||||||
Db   359  AAAAAAAAAAA  370

RESULT 12
ATAJ0058
LOCUS       ATAJ0058       2638 bp    mRNA            PLN       06-APR-1998
DEFINITION  Arabidopsis thaliana mRNA for MCM3 homolog.
ACCESSION   AJ000058
NID         g3036818
VERSION     AJ000058.1   GI:3036818
KEYWORDS    MCM3 homolog.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2638)
AUTHORS     Dodeman.V.L.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Dodeman V.L., CNRS/ERS 569, Institut de
            Biotechnologie des Plantes, Universite de Paris-Sud Bat 630, 91405
            Orsay CEDEX, FRANCE
REFERENCE   2  (bases 1 to 2638)
AUTHORS     Dodeman,V.L., Phan,T., Sabelli,P. and Bergounioux,C.
TITLE       Expression analysis of Arabidopsis thaliana MCM3 homolog during the
            cell cycle
JOURNAL     Unpublished
FEATURES             Location/Qualifiers
     source          1..2638
```

```
CDS             53..2383
                /codon_start=1
                /product="MCM3 homolog"
                /protein_id="CAA03887.1"
                /db_xref="PID:e1286040"
                /db_xref="PID:g3036819"
                /db_xref="GI:3036819"
                /db_xref="SPTREMBL:065400"
                /translation="MDVPEETRLRKRDFIQFLDSMYMEEIKALVHQKRHRLIINISD
                IHHHFRRVASRILKNPNEYMQSFCDAATEATRAIDPKYLKEGELVLGFEGYFVSRVV
                TGSVYPTPRDDKGNLLVTEYGLCKYKDHQTLSIQEVPENAAPGQLPRSVDVIAEDDIVD
                SCKPGDRVSVFGIYKALPGKSGKGSVNGVFRTILIANNIALLNKEANAPIYTKQDLDNI
                KNIARRDDAFDLLARSLAPSIYGHAWIKKAVVLLMLGGVEKNLKNGTHLRGDINMMMV
                GDPSVAKSQLLRAIMNIAPLAISTTGRGSSGVGLTAAVTSDQETGERRLEAGAMVLAD
                KGIVCIDEFDKMNDQDRVAIHEVMEQQTVIAKAGIHASLNARCSVVAANPIYGTYD
                RSLTPTKNIGLPDSLLSRFDLLFIVLDQMDAGIDSMISEHVLRMHRYKNDRSEAGPDG
                SLPYAREDNAESEMFVKYNQTLHGKKKRGQTHDKTLTIKFLKKYIHYAKHRITPKLJTD
                EASERIAEAYADLRNAGSDTKTGGLPITARTLETIIRLATAHAKMKLSSEVTKADAE
                AALKLMNFAIYHQELTEMDDREQEERQREQAEQERTPSGRGNQRRNNEDGAENDTAN
                VDSETADPMEVPDEVFGFSGTVSAARIETEFRVFGQHMRTHRLDDISIADIETVNN
                NGVGASRYSADEIMALLEKLQDDNKVMISDGKVHII"
BASE COUNT    887 a     486 c     586 g     679 t
ORIGIN

Query Match          2.7%;   Score 91.4; DB 7;   Length 2638;
Best Local Similarity  73.0%;  Pred. No. 3e-05;
Matches 116;  Conservative   1;  Mismatches  42;  Indels   0;  Gaps   0;

Qy  3252  ATATTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGTGTGACAAAATTAAAGGC  3311
          || ||||| ||| |||||||| | ||| | |||||||| ||||| || ||||| ||||||
Db  2475  ATTTTGTAAAATCTTTATTCATACCGTTATGTTTTTGCATTTACTAGTGAAGTAAAACTG  2534

Qy  3312  TTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db  2535  ATGATAAAATGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2594

Qy  3372  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAA  3410
          |||||||||||||||||||||||||||||| |||||||||||
Db  2595  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2633

RESULT 13
HSGLCNACT
LOCUS       HSGLCNACT      1887 bp    mRNA            PRI       01-JUL-1997
DEFINITION  H.sapiens mRNA for GlcNac-1-P transferase.
ACCESSION   Z82022
NID         g2239118
VERSION     Z82022.1   GI:2239118
KEYWORDS    GlcNac-1-P transferase.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1887)
AUTHORS     Eckert.V., Mazhari-Tabrizi,R., Blank,M., Mumberg,D., Funk,M. and
            Schwarz,R.
TITLE       Cloning and functional expression of the human GlcNac-1-P
            transferase, the enzyme for the committed step of the
            dolichol-cycle by heterologous complementation in yeast
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 1887)
AUTHORS     Eckert,V.
TITLE       Direct Submission
JOURNAL     Submitted (08-NOV-1996) Eckert V., University of Marburg,
            Medizinisches Zentrum fuer Hygiene, Robert-Koch-Str. 17, Marburg,
            Germany, D-35037
```

```
                /organism="Arabidopsis thaliana"
                /strain="C0"
                /sub_species="Columbia"
                /db_xref="taxon:3702"
                /tissue_type="seedling"
```

```
FEATURES       Location/Qualifiers
source         1..1887
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="hsalg7"
               /tissue_type="lung"
               /cell_type="fibroblast"
               /clone_lib="cDNA library (pRS415-vector)"
CDS            104..1306
               /codon_start=1
               /product="GlcNac-1-p transferase"
               /protein_id="CAB04787.1"
               /db_xref="PID:e280927"
               /db_xref="PID:g2239119"
               /db_xref="GI:2239119"
               /db_xref="SPTREMBL:O15216"
               /translation="MPLLINLIVSLLGFVATVTLIPAFLGHFIAARLCGQDLNKTSRQ
               QIPESQGVISAGAVFLILLFCFIPFPFLNCFVKEQCKAFPHHEFVALIGALLAICCMIF
               LGFADDVLNLRWRHKLLLPTAASLPLLMVYFTNFGNTTIVVPKPFRPIGLHLDLGIL
               YYVVMGLLAVFCTNAINILAGINGLEAGQSLVISASIVFNLVELEGDCRDDHVFSLY
               FMIPFFTTLGLLYHNWYPSRVFVGDTFCYFAGMTFAVVGILGHFSKTMLLFFMPQVF
               NFLYSLPQLLHIIPCPRHRIPRLNIKTGKLEMSYSKPKTKSLSFLGTFILKVAESLQL
               VTVHQSETEDGEFTECNNMTLINLLKVLGPIHERNLTLLLLLQIIGSAITFSIRYQ
               LVRLFYDV"
BASE COUNT     465 a       514 c      372 g      536 t
ORIGIN

Query Match
Best Local Similarity      2.7%;     Score 91.2; DB 10;  Length 1887;
Matches       99; Conservative      0; Mismatches    1; Indels    0; Gaps    0;

Qy   3297 GACAAAATTAAAGGCTTTCTATATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3356
          ||| |||| |||||  ||||  ||||||| |||||||||||||||||||||||||||||||
Db   1767 GAGACAATAAAGTGTTGTCTTTTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1826

Qy   3357 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
          |||||||||||:|||||||:||||||||||||||||||||| ||||||||||||
Db   1827 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1880

RESULT 14
AF118274
LOCUS       AF118274      1748 bp    mRNA            PRI       25-APR-1999
DEFINITION  Homo sapiens DNb-5 mRNA, partial cds.
ACCESSION   AF118274
NID         g4680228
VERSION     AF118274.1    GI:4680228
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1748)
  AUTHORS   Amler,L.C., Bauer,A., Corvi,R., Dihlmann,S., Praml,C.,
            Cavenee,W.K., Schwab,M. and Hampton,G.M.
  TITLE     Identification and Characterization of Novel Genes Located at the
            t(1;15)(p36.2)(q24) Translocation Breakpoint in the Neuroblastoma
            Cell Line NGP
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1748)
  AUTHORS   Amler,L.C. and Hampton,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-1999) Genos Biosciences, 11099 North Torrey Pines
            Road, La Jolla, CA 92037, USA
FEATURES       Location/Qualifiers
source         1..1748
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="1"
               /map="1p36.1-p36.2"
               /note="located close to the reciprocal translocation
               (1:15) (36.2:q24) in neuroblastoma cell line NGP"
```

```
CDS            <1..1602
               /note="deleted in neuroblastoma."
               /codon_start=1
               /product="DNb-5"
               /protein_id="AAD27583.1"
               /db_xref="PID:g4680229"
               /db_xref="GI:4680229"
               /translation="KNGPTAFGVLLPXXGTQTLSLSPXFGGGFGYVLGGIHWDKTGFG
               RALGGQFRVIXLFTAVTLSVTTVLTLVSIPERPLRPPSEKRAAMKSPSLPLPPSPVL
               PEEGPGDSLPSHTATNFSSPISPPSPLTPKYGSFISRDSSLTGISEFASSFGTANIDS
               VLIDCFTGHDSTLAIPGSVPRPPISVSFPRAPDGFYRQDRGLLEGREGALTSCDGD
               ILRVGSLDTSKPRSSGLLKRPQTLAIPDAAGGGPETSRRRNVTFSQVANILLNGVK
               YESELTGSSERAEQPLSVGRLCSTICNMPKALRTLCVNHFLGWLSFEGMLLFYTDFMG
               EVVFQGDPKAPHTSEAYQKYNSGVTMGCWGMCIYAISAAFYSAALEKLEEFLSVRTLY
               FIAYLAFGLGTGLATLSRNLYVVLSLCITYGILFSTLCTLPYSLLCDYYQSKKFAGSS
               ADGTRRGMGVDISLLSCQYFLAQILVSLVLVGPLTSAVGSANGVMYFSSLVSFLGCLYS
               SLFVIYEIPPSDAADEEHRPLLLNV"
BASE COUNT     318 a       578 c      504 g      341 t       7 others
ORIGIN

Query Match
Best Local Similarity     2.7%;      Score 91;  DB 42;  Length 1748;
Matches      220; Conservative      0; Mismatches  215; Indels   0; Gaps    0;

Qy   1075 CTGCGCATGCCCCGCACCCTGCCCGGCTCTTCGTGCTGGTGAGCTGCAGCTGCGATGGC 1134
          ||||||||||||||||  ||| |  | ||| |||   |||  |||  |    || |||||
Db    903 CTGCAACATGCCCAAGGCGCTACCGCACCCTGCGTCACCACTTCCTGGGGTGGCCTCTC 962

Qy   1135 ACTCATGACCTTCACGCGTGTTTACACGGATTTCGTGGCGAGGGCGTACCAGGGGCGT 1194
          ||  ||||| || |||| |  ||| |||||| |||| ||| ||    ||| |||||||
Db    963 ATTCGAGGGGATGTGCTCTTCTACACAGACTTCACGGGGCGGTGTTTCAGGGGGA 1022

Qy   1195 GCCCAGAGCTGAGCCGGGCACCGAGGCCCCGGAGACACTATGATGAAGGCGTTCGGATGGG 1254
          ||||  |||| ||  |||||  ||||| |||||  ||||| |  ||| |   |||||||||
Db   1023 CCCCAAGGCCCCGCCACCATCAGAGGCGTATCAGAGGTACAACAGCGGCGTGACCATGGG 1082

Qy   1255 CAGCCTGGGCTGTTCCTGCAGTGCGCCATCTCCGGTCTTCCTCGTCATGGACCG 1314
          |  ||| || || ||| ||| |  | |||   ||| |  | ||||||||||||
Db   1083 CTGCTGGGGCATGTGATCGTATCCAGTGCGTCGCCTTCAGTGCTCAGCTATCCTGGAGAA 1142

Qy   1315 GCTGGTGCAGCGATTCGGCACTCGAGCACTGCTATTGGCCAGTGTGGCAGCTTTCCCTGT 1374
          |||||| ||  |     ||| |  || || |||| |||  ||| ||| |||  |||||
Db   1143 GCTGGAGAGTTCCTCAGCGTCCGCACCCTACTTCATCGCCATCTCCGCCTTCGGCCT 1202

Qy   1375 GGCTGCCGGTGCCACATGCCGTTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCT 1434
          || |||||  |||||||||    | ||||||||    |||  |   |  ||||||||||
Db   1203 GGCTGGTCCGTGGCCACATGGTCGCTGGTGTCCATGTGGCCGTGGTGACAGCTTCAGCCGCCCT 1262

Qy   1435 CACCGGGTTCACCTTCTCAGCCTGCGCAGATCCTGCCCCCTACACACTGGCCTCCCCTCACCA 1494
          |  | | | |||| ||  |  ||| ||||| ||||  | ||| |||| ||||  |||||||
Db   1263 CTACGGGATTTATTTTCCACCCTGCGCACCTGCCACCTTGCCTCCTGCCTGCTCTGCGATTACTA 1322

Qy   1495 CCGGGAGAAGCAGT 1509
          | ||| || ||||
Db   1323 TCAGTAAGAAGT 1337
```

```
RESULT 15
A58523/c
LOCUS       A58523      13414 bp    DNA    circular    PAT       05-MAR-1998
DEFINITION  Sequence 27 from Patent WO9638555.
ACCESSION   A58523
NID         g3714139
VERSION     A58523.1    GI:3714139
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 13414)
  AUTHORS   Bogaert,T., Stringham,E. and Vandekerckhove,J.
  TITLE     PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH CONTROL CELL
            BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND PHARMACEUTICAL COMPOSITIONS
```

```
JOURNAL     CONTAINING THEM AND THEIR USE IN THE CONTROL OF CELL BEHAVIOUR
            Patent: WO 9638555-A 27 05-DEC-1996;
            BOGAERT THIERRY (BE)
COMMENT     Other publication AU 6123496 961218.
FEATURES             Location/Qualifiers
     source          1..13414
                     /organism="unidentified"
                     /db_xref="taxon:32644"        1 others
BASE COUNT    3854 a   3072 c   2884 g   3603 t
ORIGIN
```

```
Query Match            2.7%;  Score 90.8;  DB 5;  Length 13414;
Best Local Similarity  87.5%;  Pred. No. 3.3e-05;
Matches   98;  Conservative   1;  Mismatches  13;  Indels   0;  Gaps   0;

Qy   3299 CAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3358
          ||| ||  ||   |||| | |  ||| | | ||||||||||||||||||||||||||||||
Db  11234 CAACATCTACTTCTGTATTATTTCTATGAAAAAAAAAAAAAAAAAAAAAAAAAA 11175

Qy   3359 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
          |||||||||||:||||||||||||||||||| ||||||||||||
Db  11174 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 11123
```

Search completed: September 28, 1999, 12:21:27
Job time: 4967 sec

Copyright (c) 1993 - 1998 Compugen Ltd.

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 10:59:40 ; Search time 289.74 Seconds
                                     (without alignments)
                                     2944.554 Million cell updates/sec

Title: US-09-030-606-110
Perfect score: 3410
Sequence: 1 GGGAACCAGCCTGCACGCGC........AAAAAATAAAAAAAAAAA 3410

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database :   N_Geneseq_36:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3409.6	100.0	3410	1	V58586	Prostate tumour sp
2	3409.6	100.0	3410	1	V61201	Full length cDNA s
3	2136.4	62.7	2152	1	V71181	Consensus sequence
4	2114.8	62.0	2143	1	V71180	Clone 1711346.H, t
5	673.4	19.7	789	1	V58487	3' fragment of pro
6	673.4	19.7	789	1	V61144	3' cDNA sequence o
7	603	17.7	772	1	V58488	5' fragment of pro
c 8	603	17.7	772	1	V61145	5' cDNA sequence o
c 9	405.8	11.9	435	1	X40506	Human secreted pro
10	319.8	9.4	342	1	V71173	PS108 gene-specifi
11	291	8.5	294	1	V71174	PS108 gene-specifi
12	287	8.4	288	1	V71175	PS108 gene-specifi
13	270	7.9	272	1	V71176	PS108 gene-specifi
14	265	7.8	265	1	V71177	PS108 gene-specifi
15	256.4	7.5	258	1	V71166	PS108 gene-specifi
16	254	7.4	255	1	V71168	PS108 gene-specifi
17	247	7.2	247	1	V71169	PS108 gene-specifi
18	220	6.5	231	1	V71170	PS108 gene-specifi
19	217	6.4	217	1	V71167	PS108 gene-specifi
20	213	6.2	213	1	V71179	PS108 gene-specifi
21	210	6.2	223	1	V71172	PS108 gene-specifi
22	183	5.4	195	1	V71171	PS108 gene-specifi
23	149	4.4	151	1	V71178	PS108 gene-specifi
24	97.4	2.9	1582	1	T18831	Human survival mot
25	97.4	2.9	1582	1	T28259	Survival motor neu
26	95.2	2.8	1376	1	X04325	Human secreted pro
27	94.6	2.8	2676	1	Q02819	cDNA sequence enco
28	94.6	2.8	2671	1	Q03303	Entire porcine tra
29	94.4	2.8	2669	1	Q56925	Pig TGF-beta-3. Nu
30	94.4	2.8	2447	1	V54587	Human secretory pr
31	94.4	2.8	3899	1	V63192	cDNA from clone dt
32	93.2	2.7	1834	1	Q04690	Encodes Mammalian
33	93	2.7	2427	1	Q04107	Human pro-urokinas
34	92.6	2.7	2123	1	T59701	PTH-like peptide D
35	92.2	2.7	6020	1	Q06648	Plasminogen gene f
36	92.2	2.7	6010	1	Q11998	Human plasminogen
37	91.6	2.7	921	1	V40521	Homo sapiens CH27_
38	91.4	2.7	1066	1	Q49943	Human anti-HBs lig
39	91.4	2.7	7753	1	Q56826	Norwalk virus stra
40	90.8	2.7	6671	1	Q97228	Human DRADA gene 6
c 41	90.8	2.7	13414	1	T71321	Plasmid pCB50 enco
c 42	90.8	2.7	10288	1	T71322	Plasmid pCB51 enco
43	90.8	2.7	6671	1	T89799	Human double stran
44	90.8	2.7	6671	1	V27063	Human double-stran
45	90.8	2.7	3213	1	V55742	Human secreted pro

ALIGNMENTS

RESULT 1
V58586
ID   V58586 standard; cDNA; 3410 BP.
AC   V58586;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone L1-12.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   284..1945
FT   /*tag= a
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-805596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   P-PSDB; W69385.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 84-85; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 3410 BP; 667 A; 1014 C; 945 G; 783 T;

Query Match   100.0%; Score 3409.6; DB 1; Length 3410;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 GGGAACCAGCCTGCACGCGCTGGCTCCGGGTGACAGCCGGCGCCTCGGCCAGGATCTGA 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGAACCAGCCTGCACGCGCTGGCTCCGGGTGACAGCCGGCGCCTCGGCCAGGATCTGA 60

Qy   61 GTGATGAGACGTGTCCCCACTGAGGTGCCCCACACAGCAGGTGTTGAGCATGGGCTGAG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTGATGAGACGTGTCCCCACTGAGGTGCCCCACACAGCAGGTGTTGAGCATGGGCTGAG 120

Qy 121 AAGCTGGACCCGCACCAAAGGGCTGGCAGAAATGGGCGCCTCGGCTGATTCCTAGGCAGTT 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AAGCTGGACCCGCACCAAAGGGCTGGCAGAAATGGGCGCCTCGGCTGATTCCTAGGCAGTT 180

Qy 181 GGCGGCAGCAAGGAGGAGAGGCCCAGCTTCTGGAGCAGAAGCCAGAGCAGCAGTTCTG 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GGCGGCAGCAAGGAGGAGAGGCCCAGCTTCTGGAGCAGAAGCCAGAGCAGCAGTTCTG 240

Qy 241 GAGTGCCTGAACGGCCCCTGAGCCTCAGCCCTACCCGCCGGCCCTGGCCCACTATGGTCCAGAGGCTGTG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GAGTGCCTGAACGGCCCCTGAGCCTCAGCCCTACCCGCCGGCCCTGGCCCACTATGGTCCAGAGGCTGTG 300

Qy 301 GGTGAGCGCCGTGCTGCGGCACCGGAAAGCCCAGCTTGTGGTCAACCTGCTAACCTT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GGTGAGCGCCGTGCTGCGGCACCGGAAAGCCCAGCTTGTGGTCAACCTGCTAACCTT 360

Qy 361 TGGCCTGGAGGTGTGTTTGCCGCAGGCATCACCTATGTGCCGCCCTCTGCTGCTGGAAGT 420

```
Db  361  ||||||||||||||||||||
Qy  361  TGGCCTTGGAGCTGTGTTGTTTGGCCGGCATCCAGGCCATGGTGCTGGGCATTGGTCCAGTTGCTGGGCCTGGT  420
Db  421  ||||||||||||||||||||
Qy  421  GGGGGTAGAGGAGGAAGTTCATGACCAGTTCATGAGCCTGGAGCGTCCCGTCCCGCCGG  480
Db  481  ||||||||||||||||||||
Qy  481  CTGTGTCCCGCTCCTAGGCTCCAGTGACCACTGGCCGTGGAGGCTATGGCCCGCCG  540
Db  541  ||||||||||||||||||||
Qy  541  GCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTCTCATCCCAAGGGC  600
Db  601  ||||||||||||||||||||
Qy  601  CGGCTGGCTAGCAGGGCTGCTGTGCGCCCAGGATCCCAGGCCCCTGGAGCTGGCACTCCACTGGAGGGCCCCT  660
Db  661  ||||||||||||||||||||
Qy  661  CCTGGGCCGTGGGGCTGCTGGGACTTCGTGGCCAGGTGTGCTTCACTCCACTGGAGGCCCCT  720
Db  721  ||||||||||||||||||||
Qy  721  GCTCTCTGACCTCTTCCGGGACCCGGACCACTGTCGCGGACCCTACTGTGTCTATGCCTT  780
Db  781  ||||||||||||||||||||
Qy  781  CATGGATCAGTCTTGGGGGCTGCCTGGGCTACCTCCTGCCATTGACTGGGACACCAG  840
Db  841  ||||||||||||||||||||
Qy  841  TGCCCTGGCCCCCTACCTGGGCACCCAGGAGGAGTGCCTCTTTGCCTCTTTGGCCTCTCACCCTCAT  900
Db  901  ||||||||||||||||||||
Qy  901  CTTCCTCACCTGCTAGCAGCCACCACTGCTGGTGGCTAGCAGCCAGCAGCCGCCTGGGCCCCCAC  960
Db  961  ||||||||||||||||||||
Qy  961  CGAGCCAGCAGAGAGGGGCTGTCGGCCCCCTACCTGGGCGCCGCGCTCTTGCGCCTCTTCGCGCCGGGGC  1020
Db 1021  ||||||||||||||||||||
Qy 1021  CCGCTTGGCTTTCCGGAACCTGGCCGCGCCTTGGGCGCCCGGCTCTTCCCCGGCGTGCTGCGCG  1080
Db 1081  ||||||||||||||||||||
Qy 1081  CATGCGGCCGGGGCACCCGCACCCGGGCAGGCCGGGCTCTTCGCGGCTGTGGGGCGTGCCGCG  1140
Db 1141  ||||||||||||||||||||
Qy 1141  GACCTTCACGCTGTTTCCGGAGCTTTACACGGATTCGTGGGGCGACACTATGATGAAGGCGTTCGGATGGCACTCAT  1200
Db 1201  ||||||||||||||||||||
Qy 1201  AGCTGAGCCGGGCCAGGGCACCGAGCCCCGGAGACACTATGATGAAGGCGTTCGGATGGCAGCCT  1260
Db 1261  ||||||||||||||||||||
Qy 1261  GGGGCTGTTCCTGCAGTGCGCCATCTCCCGGTCTTCTCTGGTCCAGTCGCAGTC  1320
Db 1321  ||||||||||||||||||||
Qy 1321  GCAGCGGATTCGGCCACTCGAGCAGTCTATTTGGCCGTGTGGCCGTGTCCACAGTCGGCCTGCTGTCCCACAGT  1380
Db 1381  ||||||||||||||||||||
Qy 1381  CGGTGCCACATGCCGTCCCACAGTGTGGCCGTGTCGCCGTCGCCTACACACTTGCCTCCGTGTTCCTCCACCGG  1440
Db 1441  ||||||||||||||||||||
Qy 1441  GTTCACCTTCTCAGCCCTGCAGATCCCTGCCTACACACTGGCCTCCCTCTACCACCGGGA  1500
```

```
Db 1441  ||||||||||||||||||||
Qy 1441  GTTCACCTTCTCAGCCCTGCAGATCCTGCCTACACACTGGCCTCCCTCTACCACCGGGA  1500
Db 1501  ||||||||||||||||||||
Qy 1501  GAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGAGGTGCTAGCAGTGAGGACAGT  1560
Db 1501  GAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGAGGTGCTAGCAGTGAGGACAG  1560
Db 1561  ||||||||||||||||||||
Qy 1561  CCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCTAATGGACACGT  1620
Db 1561  CCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCTAATGGACACGT  1620
Db 1621  ||||||||||||||||||||
Qy 1621  GGGTGCTGGAGGCAGTGGCCTGCTCCCACCCGCCGGCGCCTCTGCGGGGGCCTCTGCCCTG  1680
Db 1621  GGGTGCTGGAGGCAGTGGCCTGCTCCCACCCGCCGGCGCCTCTGCGGGGGCCTCTGCCCTG  1680
Db 1681  ||||||||||||||||||||
Qy 1681  TGATGTCTCCGTACGTGTGGTGGGTGGGTGGTGCGGGCACGCCCAGGGTGGTTCCGGGCCG  1740
Db 1681  TGATGTCTCCGTACGTGTGGTGGGTGGGTGGTGCGGGCACGCCCAGGGTGGTTCCGGGCCG  1740
Db 1741  ||||||||||||||||||||
Qy 1741  GGGCATCTGCCTGGACCTGGCCATCCTGGATAGTGCCTTCGCTCCGTGTCCCAGTGGCCCCC  1800
Db 1741  GGGCATCTGCCTGGACCTGGCCATCCTGGATAGTGCCTTCGCTCCGTGTCCCAGTGGCCCCC  1800
Db 1801  ||||||||||||||||||||
Qy 1801  ATCCCTGTTTATGGGCCTCCAGTCTGTCCAGGCCCAGTCGCTGTCACTGCCTATATGGTGTC  1860
Db 1801  ATCCCTGTTTATGGGCCTCCAGTCTGTCCAGGCCCAGTCGCTGTCACTGCCTATATGGTGTC  1860
Db 1861  ||||||||||||||||||||
Qy 1861  TGCCGCAGGCCTGGGTCTGTGCGCCATTACTTGCTGCCGCCATTACTTGCTACACAGGTAGTATTTGACAAGAG  1920
Db 1861  TGCCGCAGGCCTGGGTCTGTGCGCCATTACTTGCTGCCGCCATTACTTGCTACACAGGTAGTATTTGACAAGAG  1920
Db 1921  ||||||||||||||||||||
Qy 1921  CGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCT  1980
Db 1921  CGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCT  1980
Db 1981  ||||||||||||||||||||
Qy 1981  CACTGGTCCCAGCTCCCCGTCTCCCGCTCCTCGTTGAGCCCCATGGGCGTGCCGGCTGCCGCCAGT  2040
Db 1981  CACTGGTCCCAGCTCCCCGTCTCCCGCTCCTCGTTGAGCCCCATGGGCGTGCCGGCTGCCGCCAGT  2040
Db 2041  ||||||||||||||||||||
Qy 2041  TTCTGTTGCTGCCAAAGTAATGTGCTCTGTGCTCTGCTGCCACCCTGTGCTGCCGAGGTGCGTA  2100
Db 2041  TTCTGTTGCTGCCAAAGTAATGTGCTCTGTGCTCTGCTGCCACCCTGTGCTGCCGAGGTGCGTA  2100
Db 2101  ||||||||||||||||||||
Qy 2101  GCTGCACAGCTGGGGCTGGGGCTGGGGGTTCCAGTCTGGACTTATACAGGGAGGCCAGAAGGGCTCC  2160
Db 2101  GCTGCACAGCTGGGGCTGGGGCTGGGGGTTCCAGTCTGGACTTATACAGGGAGGCCAGAAGGGCTCC  2160
Db 2161  ||||||||||||||||||||
Qy 2161  ACTGGAGGCCTTCCAAGGGGTTTCAGTCTGCAGGTGGATTCTGGACTTATACAGGGAGGCTGAATAAACTCAGTCACCTG  2220
Db 2161  ACTGGAGGCCTTCCAAGGGGTTTCAGTCTGCAGGTGGATTCTGGACTTATACAGGGAGGCTGAATAAACTCAGTCACCTG  2220
Db 2221  ||||||||||||||||||||
Qy 2221  ATGCACTGGAATGCGGGGACTCTGCAGGTTAACCTGGTTAATGTAGCTCTTGCATGGGAG  2280
Db 2221  ATGCACTGGAATGCGGGGACTCTGCAGGTTAACCTGGTTAATGTAGCTCTTGCATGGGAG  2280
Db 2281  ||||||||||||||||||||
Qy 2281  CTCCTAGTTGAGACACCTAGAGAAGGGTTTTTGGGAGCTGAATAAACTCAGTCACCTG  2340
Db 2281  CTCCTAGTTGAGACACCTAGAGAAGGGTTTTTGGGAGCTGAATAAACTCAGTCACCTG  2340
Db 2341  ||||||||||||||||||||
Qy 2341  GTTTCCCATCTCTAAGCCCTTAACCTGCGTTAATGAACATATGACTTATTGTAGGGGAAGA  2400
Db 2341  GTTTCCCATCTCTAAGCCCTTAACCTGCGTTAATGAACATATGACTTATTGTAGGGGAAGA  2400
Db 2401  ||||||||||||||||||||
Qy 2401  TTTCTAGGATGATGAAAACACTCCTCCCACAAGACCCTTTTTGCT  2460
Db 2401  TTTCTAGGATGATGAAAACACTCCTCCCACAAGACCCTTTTTGCT  2460
Db 2461  ||||||||||||||||||||
Qy 2461  GTCCTGAGGGGCAACACACAGAACCAGTCCCCTGTCTTTTTTGCT  2520
Db 2461  GTCCTGAGGGGCAACACACAGAACCAGTCCCCTGTCTTTTTTGCT  2520
Db 2521  ||||||||||||||||||||
Qy 2521  GATCCACCCCCTCTTACCTTTATCAGGATGTGGCCTGTTGTGTTCTGTTGCCATCA  2580
Db 2521  GATCCACCCCCTCTTACCTTTATCAGGATGTGGCCTGTTGTGTTCTGTTGCCATCA  2580
```

```
PI   Dillon DC, Xu J;
DR   WPI; 98-609986/51.
DR   P-PSDB; W71869.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 79-80; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence   3410 BP;   667 A;   1014 C;   945 G;   783 T;

     Query Match         100.0%;   Score 3409.6;   DB 1;   Length 3410;
     Best Local Similarity  100.0%;   Pred. No. 0;
     Matches 3410;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy      1 GGGAACCAGCCTGCACGCGCTGCTCCGGGTGACAGCCGGCGCCCTCGCCCTCAGGATCTGA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGAACCAGCCTGCACGCGCTGCTCCGGGTGACAGCCGGCGCCCTCGCCCTCAGGATCTGA 60

Qy     61 GTGATGAGACGTGTCCCCACTGAGGTGCCCACCACAGCAGCAGGTGTTGAGCATGGCGTGAG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTGATGAGACGTGTCCCCACTGAGGTGCCCACCACAGCAGCAGGTGTTGAGCATGGCGTGAG 120

Qy    121 AAGCTGACCGGCACCAAAGGGCTGCAGAAATGGGCGCCTGGCTGATTCCTAGGCAGTT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGCTGACCGGCACCAAAGGGCTGCAGAAATGGGCGCCTGGCTGATTCCTAGGCAGTT 180

Qy    181 GGCGGCAGCAAGGAGGAGGAGGCCGCAGCTTCTGGAGCAGAGCGAGCAGCAGTTCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCGGCAGCAAGGAGGAGGAGGCCGCAGCTTCTGGAGCAGAGCGAGCAGCAGTTCTG 240

Qy    241 GAGTGCCTGAACGGCCGCCCTGAGCCCACCGGCACCATGGTCTCCAGAGGCGTGTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGTGCCTGAACGGCCGCCCTGAGCCCACCGGCACCATGGTCTCCAGAGGCGTGTG 300

Qy    301 GGTGAGCCGCCGCTGCTGCGGCGAAAGCCCAGCTCTTGGTCAACCTGCTCTAACCTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGTGAGCCGCCGCTGCTGCGGCGAAAGCCCAGCTCTTGGTCAACCTGCTCTAACCTT 360

Qy    361 TGGCCTGGAGGTGTGTTGGGCCGCATCACCTATGTGCCGCCTCTGCTGCTGGAAGT 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGCCTGGAGGTGTGTTGGGCCGCATCACCTATGTGCCGCCTCTGCTGCTGGAAGT 420

Qy    421 GGGGGTAGAGGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTCGGCCTGGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGGGGTAGAGGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTCGGCCTGGT 480

Qy    481 CTGTGTCCCGCTCCTAGGCTCCAGTGACCAGTGACGCGTATGGCGCGCCGCCG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGTGTCCCGCTCCTAGGCTCCAGTGACCAGTGACGCGTATGGCGCGCCGCCG 540

Qy    541 GCCCTTCATCTGGCACTGGCACTGCTGCTCTGGGAGCCTCTTCTCCATCCAAGGGC 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCCCTTCATCTGGCACTGGCACTGCTGCTCTGGGAGCCTCTTCTCCATCCAAGGGC 600

Qy    601 CGGCTGGCTAGCAGGGCTGCTGCTGCCCAGGCCCCCTGGAGCTGGCACTGCTCAT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CGGCTGGCTAGCAGGGCTGCTGCTGCCCAGGCCCCCTGGAGCTGGCACTGCTCAT 660

Qy    661 CCTGGGCGTGGGCTGCTGGGACTTCTGTGGCCAGGTGCTTCACTCCACTGGAGGCCCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCTGGGCGTGGGCTGCTGGGACTTCTGTGGCCAGGTGCTTCACTCCACTGGAGGCCCT 720

Qy    721 GCTCTCTGACCTCTTCCGGGACCCGACCACTCGCCGCCAGCCTGTCTGTCTATGCCTT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCTCTCTGACCTCTTCCGGGACCCGACCACTCGCCGCCAGCCTGTCTGTCTATGCCTT 780

Qy    781 CATGATCAGTCTTGGGGCTGCTGGGCTACCTCCTCGCTGCCATTGACTGGGACACCAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   2581 CAGAGACACAGGCATTTAAATATTTAACTTATTTATTTAACAAAGTAGAAGGGAATCCAT 2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 CAGAGACACAGGCATTTAAATATTTAACTTATTTATTTAACAAAGTAGAAGGGAATCCAT 2640

Qy   2641 TGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTGGGGGATCCCCAACAATCA 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 TGCTAGCTTTTCTGTTGGTGTCTAATATTTGGGTAGGGTGGGGGATCCCCAACAATCA 2700

Qy   2701 GGTCCCCTGAGATAGCTGGTCATTGGGCTGATCATTTCTTCTCCTGGGGT 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 GGTCCCCTGAGATAGCTGGTCATTGGGCTGATCATTTCTTCTCCTGGGGT 2760

Qy   2761 CTGGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTCACTCATCCATCCAAATGATAAT 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 CTGGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTCACTCATCCATCCAAATGATAAT 2820

Qy   2821 TCCAAATGCTGTTACCCAGGTTAGGGTGTTGAAGGAAGGTAGAGGGCTTCAGGT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 TCCAAATGCTGTTACCCAGGTTAGGGTGTTGAAGGAAGGTAGAGGGCTTCAGGT 2880

Qy   2881 CTCAACGGCTTCCCTAACCACCCCTTCCTTCTTGGCCCAGCCTGGTCCCCCACTTCCA 2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 CTCAACGGCTTCCCTAACCACCCCTTCCTTCTTGGCCCAGCCTGGTCCCCCACTTCCA 2940

Qy   2941 CTCCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTCCCCTACC 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CTCCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTCCCCTACC 3000

Qy   3001 CCCAACTTTCCCCTACCCCCCAACTTCCCACCAGTCCCACAACCCTGTTTGGGAGCTACT 3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CCCAACTTTCCCCTACCCCCCAACTTCCCACCAGTCCCACAACCCTGTTTGGGAGCTACT 3060

Qy   3061 GCAGGACCAGAAGCACAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCCAGAGT 3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GCAGGACCAGAAGCACAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCCAGAGT 3120

Qy   3121 ATATCTGTGCTTGGGGAATCTCACACAGAAAACTCAGGAGCACCCCCTGCCCTGAGCGTAAGG 3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 ATATCTGTGCTTGGGGAATCTCACACAGAAAACTCAGGAGCACCCCCTGCCCTGAGCGTAAGG 3180

Qy   3181 GAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGGTCTTATTTATT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGGTCTTATTTATT 3240

Qy   3241 TAGGCGGGGTGAATATTTTATACTGTAAGTGCAATCAGAGTATAATGTTTATGGTGACA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TAGGCGGGGTGAATATTTTATACTGTAAGTGCAATCAGAGTATAATGTTTATGGTGACA 3300

Qy   3301 AAATTAAAGGCTTTCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 AAATTAAAGGCTTTCTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360

Db   3361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAA 3410
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAA 3410
```

```
RESULT  2
V61201
ID   V61201 standard; cDNA; 3410 BP.
AC   V61201;
DT   06-JAN-1999 (first entry)
DE   Full length cDNA sequence of prostate tumour clone L1-12.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   25-FEB-1998; US-020956.
PR   09-FEB-1998; US-805099.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
```

```
Db   781 CATGATCAGTCTTGGGGTCCTGGCTACCTCCTGCCATTGACTGGGACACCAG 840

Qy   841 TGCCCTGGCTGCCCCCTACCTGGCACCCAGGAGGAGTGCCTCTTTGCCTGCTCACCCTCAT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 TGCCCTGGCTGCCCCCTACCTGGCACCCAGGAGGAGTGCCTCTTTGCCTGCTCACCCTCAT 900

Qy   901 CTTCCCACCTCGTAGCAGCCACACTGCTGGCTGGCTGGTGGTGAGGAGGCAGCGGCTGGGCCCCAC 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CTTCCCACCTCGTAGCAGCCACACTGCTGGCTGGCTGGTGGTGAGGAGGCAGCGGCTGGGCCCCAC 960

Qy   961 CGAGCCAGCAGAAGGGCTGTCGCCCCTCCTTGTCGCCCACTGCTGTCCATGCCGGGC 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 CGAGCCAGCAGAAGGGCTGTCGCCCCTCCTTGTCGCCCACTGCTGTCCATGCCGGGC 1020

Qy  1021 CCGCTTGGCTTCCGGAACCTGGGCGCCCTGCCCGGCGTGCACCAGCTGTGCTGCCG 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CCGCTTGGCTTCCGGAACCTGGGCGCCCTGCCCGGCGTGCACCAGCTGTGCTGCCG 1080

Qy  1081 CATGCCCCGCCACCCTGCGCGCCGGCCTCTTCGTTGCTGGTGCAGCTGGATGGCACTCAT 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 CATGCCCCGCCACCCTGCGCGCCGGCCTCTTCGTTGCTGGTGCAGCTGGATGGCACTCAT 1140

Qy  1141 GACTTCACGCGTGTTTTACACGGATTTCGTGGGCGAGGGCGTACCAGGCGTGCCCAG 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GACTTCACGCGTGTTTTACACGGATTTCGTGGGCGAGGGCGTACCAGGCGTGCCCAG 1200

Qy  1201 AGCTGAGCCGGGCACCGGGCACGGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCT 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AGCTGAGCCGGGCACCGGGCACGGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCT 1260

Qy  1261 GGGGCTGTTCCTGCAGTGCGCCATCTCCCGTCTTCTCTCCGGTCATGGACCGGCTGGT 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GGGGCTGTTCCTGCAGTGCGCCATCTCCCGTCTTCTCTCCGGTCATGGACCGGCTGGT 1320

Qy  1321 GCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCGGTGCCGTGGCAGCTTTCCCTGTGGCTGC 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 GCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCGGTGCCGTGGCAGCTTTCCCTGTGGCTGC 1380

Qy  1381 CGGTGCCACATGCCGTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCGCCGCCCTCACCGG 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CGGTGCCACATGCCGTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCGCCGCCCTCACCGG 1440

Qy  1441 GTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTACCACCGGGA 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTACCACCGGGA 1500

Qy  1501 GAAGCAGGTGTCCTGCCGCCCAAATACCGAGGGGACTGGAGGTCTAGCAGTGCGAGGACACGT 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GAAGCAGGTGTCCTGCCGCCCAAATACCGAGGGGACTGGAGGTCTAGCAGTGCGAGGACACAG 1560

Qy  1561 CCTGATGACCAGCTTCCTGCCAGCCTAAGCCTGGAGCTCCCTTCCCATGGACACGT 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 CCTGATGACCAGCTTCCTGCCAGCCTAAGCCTGGAGCTCCCTTCCCATGGACACGT 1620

Qy  1621 GGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCTCCACCGCCGCCTGCGCCTG 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GGGTGCTGGAGGCAGTGGCCTGCTCCCACCTCCACCCTCCACCGCCGCCTGCGGGCCG 1680

Qy  1681 TGATGTCTCCCGTACGTGGTGGTGCGTACTGTCACTGCCTATATGTGTC 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 TGATGTCTCCCGTACGTGGTGGTGCGTACTGTCACTGCCTATATGTGTC 1740

Qy  1741 GGGCATCGCCTCGCCACGTCGGCGACCTCGGATAGTGCCTTCCTGCGTGTCCGGTTGGCCCC 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 GGGCATCGCCTCGCCACGTCGGCGACCTCGGATAGTGCCTTCCTGCGTGTCCGGTTGGCCCC 1800

Qy  1801 ATCCCTGTTTATGGCTCCATTGTCCCGCGCCATTGTCCAGCTCAGCGTCTGTCACTGCGCTATATGCTGTC 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 ATCCCTGTTTATGGCTCCATTGTCCCGCGCCATTGTCCAGCTCAGCGTCTGTCACTGCGCTATATGCTGTC 1860

Qy  1861 TGCCGCAGGCCTGGGTGCTGGTGTCCGTGGTGGTCGCCATTTACTTTGCTCCACAGGTAGTATTTGACAAGAG 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 TGCCGCAGGCCTGGGTGCTGGTGTCCGTGGTGGTCGCCATTTACTTTGCTCCACAGGTAGTATTTGACAAGAG 1920

Qy  1921 CGACTTGGCCAAATACTCAGCGTGAGAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCCT 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 CGACTTGGCCAAATACTCAGCGTGAGAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCCT 1980

Qy  1981 CACTGGGTCCCAGCTCCCGCTCCCGCTCTTGCCTGCTCCGTGTAGCCCCATGGGGCTGCCGGCTGCCGCCCAGT 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 CACTGGGTCCCAGCTCCCGCTCCCGCTCTTGCCTGCTCCGTGTAGCCCCATGGGGCTGCCGGCTGCCGCCCAGT 2040

Qy  2041 TTCTTGTTGCTGCCAAAGTAATGTGGCTCTGCTGCCACCCTGTCTGCTGCTGAGTGGTGCGTA 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 TTCTTGTTGCTGCCAAAGTAATGTGGCTCTGCTGCCACCCTGTCTGCTGCTGAGTGGTGCGTA 2100

Qy  2101 GCTGCACAGCTGGGGGCTGGGGGCGTGGGGCGTCCCCTCCCTCTCTCCCCAGTCTCTAGGGCTGCCTG 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 GCTGCACAGCTGGGGGCTGGGGGCGTGGGGCGTCCCCTCCCTCTCTCCCCAGTCTCTAGGGCTGCCTG 2160

Qy  2161 ACTGGAGGCCTTCAAGGGGTTCAGTCTGGACTTATCAGGGAGGCCAGAAGGGCTCC 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 ACTGGAGGCCTTCAAGGGGTTCAGTCTGGACTTATCAGGGAGGCCAGAAGGGCTCC 2220

Qy  2221 ATGCACTGGAATGCCGGGACTCTGCAGTGGATTGAGTTAACCAGGTTAACAGCTAGC 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 ATGCACTGGAATGCCGGGACTCTGCAGTGGATTGAGTTAACCAGGTTAACAGCTAGC 2280

Qy  2281 CTCCTAGTTGAGACACACCTAGAGAAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 CTCCTAGTTGAGACACACCTAGAGAAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG 2340

Qy  2341 GTTTCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTCTTGCATGGGAG 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 GTTTCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTCTTGCATGGGAG 2400

Qy  2401 TTTCTAGGATGAAAACTCTCCCATGGAGTTGAACATATGACTTATTTGTAGGGGAAGA 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TTTCTAGGATGAAAACTCTCCCATGGAGTTGAACATATGACTTATTTGTAGGGGAAGA 2460

Qy  2461 GTCCTGAGGGGCAACACAGAACCAGGTCCCCTCAGCCCACAGCACTGTCTTTTTGCT 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 GTCCTGAGGGGCAACACAGAACCAGGTCCCCTCAGCCCACAGCACTGTCTTTTTGCT 2520

Qy  2521 GATCCACCCCCTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCCATCA 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 GATCCACCCCCTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCCATCA 2580

Qy  2581 CAGAGACACAGGCATTTAAATATTAAACTTAAATATTAACAAAGTAGAAGGAATCCAT 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 CAGAGACACAGGCATTTAAATATTAAACTTAAATATTAACAAAGTAGAAGGAATCCAT 2640

Qy  2641 TGCTAGCTTTCTGTGTTGGTGTCTAATATTTGGGGGATCCCAACAATCA 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 TGCTAGCTTTCTGTGTTGGTGTCTAATATTTGGGGGATCCCAACAATCA 2700

Qy  2701 GGTCCCTGAGATAGCTGGTCATTGGGCTGATCATTGCCAGAATCTTCTCTCCTGGGGT 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 GGTCCCTGAGATAGCTGGTCATTGGGCTGATCATTGCCAGAATCTTCTCTCCTGGGGT 2760

Qy  2761 CTGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAATGATAAT 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 CTGCCCCCAAAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAATGATAAT 2820

Qy  2821 TCCAAATGCTGTTACCCAAGGTAGGGTGTTGAAGGAAGGTAGAGGGCTTCAGGT 2880
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 TCCAAATGCTGTTACCCAAGGTAGGGTGTTGAAGGAAGGTAGAGGGCTTCAGGT 2880

Qy  2881 CTCAACGGCTTCCCTAACCACCCCTCTTCGGCCCAGCCGTTCCCCCACTTCAGT 2940
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881 CTCAACGGCTTCCCTAACCACCCCTCTTCGGCCCAGCCGTTCCCCCACTTCCA 2940

Qy  2941 CTCCCCTACTCTCTCTAGGACTGGCTGATGAAGGCCACTGCCCAAAATTCCCCTACC 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 CTCCCCTACTCTCTCTAGGACTGGCTGATGAGAGGCCACTGCCCAAAATTCCCCTACC 3000
```

```
Query Match             62.7%;  Score 2136.4;  DB 1;  Length 2152;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 2149;  Conservative 0;  Mismatches 0;  Indels 1;  Gaps 1;

Qy   1177 GGGGTGTACCAGGCGGTGCCAGAGCTCACCGGGACCTGAGAGTGCCGGAGCACTATGA 1235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGGTGTACCAGGCGGTGCCAGAGCTCACCGGGACCTGAGAGTGCCGGAGCACTATGA 60

Qy   1237 TGAAGGCGTTCGGATGGCAGCTCGGGCGTGTTCCTGCAGTGCCGCCATCTCCTGGTCTT 1296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGAAGGCGTTCGGATGGCAGCTCGGGCGTGTTCCTGCAGTGCCGCCATCTCCTGGTCTT 120

Qy   1297 CTCTCTGGTCATGGACCGGCTGGTGCAGCGAATTCGGCACTGCAGTCTATTTGGCCAG 1355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCTCTGGTCATGGACCGGCTGGTGCAGCGAATTCGGCACTGCAGTCTATTTGGCCAG 180

Qy   1357 TGTGCAGCTTCCCTGTGCTGCCGGTGCTGCCGGTGCACATGCCTGTCTCCCAGTGTTTATGGTGACA 1416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTGCAGCTTCCCTGTGCTGCCGGTGCTGCCGGTGCACATGCCTGTCTCCCAGTGTTTATGGTGACA 240

Qy   1417 GACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCGAGATCCGCCCTACAC 1476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACAGCTTCAGCCGCCCTCACCGGGTTCACCTTCTCAGCCCTGCGAGATCCGCCCTACAC 300

Qy   1477 ACTGGCCTCCCTCCACCACCGGGAGAAGAAGCAGCTGATGACCAGCTTCTGCCAAATACCGAGGGGACAC 1536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACTGGCCTCCCTCCACCACCGGGAGAAGAAGCAGCTGATGACCAGCTTCTGCCAAATACCGAGGGGACAC 360

Qy   1537 TGGAGGTGCTAGCAGTGAGGACGAGACCAGCCTGATGACCAGCTTCTGCCAGGCCTAAGCCTGG 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGAGGTGCTAGCAGTGAGGACGAGACCAGCCTGATGACCAGCTTCTGCCAGGCCTAAGCCTGG 420

Qy   1597 AGCTCCCTTCCCTAATGACACGTGGGTGCTGGAGGCAGTGAGGCTCAGCTCCATTGTCCAGCTCAGCCA 1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCTCCCTTCCCTAATGACACGTGGGTGCTGGAGGCAGTGAGGCTCAGCTCCATTGTCCAGCTCAGCCA 480

Qy   1657 CGGGCTCGCGGGCCTCGCTGCCTATATGTCTCCGTAGCTGTGGTGGTGGTGAGCCCAC 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGGGCTCGCGGGCCTCGCTGCCTATATGTCTCCGTAGCTGTGGTGGTGGTGAGCCCAC 540

Qy   1717 CGAGGCCAGGTGGTTCCGGGCATCTGCCGGGCATCCGGTCCCATCCTGGATAGTGC 1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CGAGGCCAGGTGGTTCCGGGCATCTGCCGGGCATCCGGTCCCATCCTGGATAGTGC 600

Qy   1777 CTTCCTGCGTGTCCCAGGTGGCCTATATTGGCTCACTGCTCCAGCTCAGCTCCA 1836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTTCCTGCGTGTCCCAGGTGGCCTATATTGGCTCACTGCTCCAGCTCAGCCA 660

Qy   1837 GTCTGTCACTGCCTATATGTCTGCTGCCCAGGCTGGGTCTGGTCGCCATTACTTTGC 1896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTCTGTCACTGCCTATATGTCTGCTGCCCAGGCTGGGTCTGGTCGCCATTACTTTGC 720

Qy   1897 TACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAG 1956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAG 780

Qy   1957 CACATTGGGGTGGAGGGCCTGCCTCACTGGGTCCCCAGCTCCCGCTCTGTTAGCCCCAT 2016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CACATTGGGGTGGAGGGCCTGCCTCACTGGGTCCCCAGCTCCCGCTCTGTTAGCCCCAT 840

Qy   2017 GGGGCTGCCGGGCTGGCCGGCCAGTTCTGCTGCTGCCAAAGTAATGTGGCTCTCTGCTGC 2076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGGGCTGCCGGGCTGGCCGGCCAGTTCTGCTGCTGCCAAAGTAATGTGGCTCTCTGCTGC 900

Qy   2077 CACCCTGTGCTGCTGAGGTGCGTAGCTGCCACAGCTGGGGCGTGGGGCTGGGCGCCTCCCTCCTCT 2136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CACCCTGTGCTGCTGAGGTGCGTAGCTGCCACAGCTGGGGCGTGGGGCTGGGCGCCTCCCTCCTCT 960

Qy   2137 CTCCCCAGTCTCTAGGGCCTCGCTGACTGACTGAGGGGTTCAGTCTGGACTT 2196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CTCCCCAGTCTCTAGGGCCTCGCTGACTGACTGAGGGGTTCAGTCTGGACTT 1020

Qy   2197 ATACAGGGAGGGCCAGAAGGGCTCCATGCACTGGAATGCGGGACTCTGCAGGTGGATTAC 2255
```

```
Qy   3001 CCCAACTTCCCTACCCCAACTTCCCCACCAGCTCCACAACCCTGTTGGAGCTACT 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 CCCAACTTCCCTACCCCAACTTCCCCACCAGCTCCACAACCCTGTTGGAGCTACT 3060

Qy   3061 GCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGT 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 GCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGT 3120

Qy   3121 ATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 ATATCTGTGCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGG 3180

Qy   3181 GAGGTCTTATCTCTCAGGGGGGTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATT 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GAGGTCTTATCTCTCAGGGGGGTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATT 3240

Qy   3241 TAGCGGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TAGCGGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300

Qy   3301 AAATTAAAGGCTTTCTTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 AAATTAAAGGCTTTCTTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360

Qy   3361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
          ||||||||||||||||||||||||||||||||||||||||||||
Db   3361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
```

```
RESULT   3
V71181
ID   V71181 standard; cDNA; 2152 BP.
AC   V71181;
DT   12-FEB-1999 (first entry)
DE   Consensus sequence of the PS108 gene derived from overlapping clones.
KW   PS108 gene: prostate disease: benign prostatic hyperplasia: BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Synthetic.
OS   Homo sapiens.
FH   Key           Location/Qualifiers
FT   CDS           2..769
FT                 /*tag= a
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
DR   P-PSDB; W85068.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1: Fig 1A-E; 122pp; English.
CC   The present sequence represents the consensus sequence for a PS108
CC   gene. The sequence is derived from overlapping clones V71166-79. The
CC   clone sequences are PS108 gene-specific. They are used in the method
CC   of the invention. The specification describes a method for detecting the
CC   presence of a target PS108 polynucleotide in a test sample. The method
CC   comprises contacting the test sample with at least 1 PS108-specific
CC   polynucleotide or complement, and detecting the presence of the target
CC   PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence   2152 BP;  419 A;  622 C;  569 G;  542 T;
```

us-09-030-606-110.rng

Wed Sep 29 14:26:55 1999

```
Db  2101 ATCAGAGTATAATGTTTATGTGTGACAAAATTAAAGGCTTTCTTATATGTTTA 2152
        |||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT  4
V71180
ID   V71180 standard; cDNA; 2143 BP.
AC   V71180;
DT   12-FEB-1999  (first entry)
DE   Clone 1711346IH, the PS108 gene contig full length sequence.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   The present sequence represents the full length contig of the PS108
CC   gene, as represented by clone 1711346IH. This clone is the contig of
CC   overlapping clones V71166-79. The clone sequences are PS108
CC   gene-specific. They are used in the method of the invention. The
CC   specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence  2143 BP;     418 A;     621 C;     563 G;     541 T;

Query Match          62.0%;   Score 2114.8;   DB 1;  Length 2143;
Best Local Similarity 99.88%; Pred. No. 0;
Matches 2139;  Conservative   0;  Mismatches   2;  Indels   3;  Gaps   2;

Qy  1185 ACCAGGGCGTGCCCAGAGCTGAGCCGGCACCGAGGCCCGGAGACACTATGATGAAGGCG 1244
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ACCAGGGCGTGCCCAGAGCTGAGCCGGCACCGAGGCCCGGAGACACTATGATGAAGGCG 60

Qy  1245 TTCGGATGGGCAGCCTGGGCGTGTTCCTGCAGTGCGCCCATCTCCCGGCTCTTCTCTGG 1304
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TTCGGATGGGCAGCCTGGGCGTGTTCCTGCAGTGCGCCCATCTCCCGGCTCTTCTCTGG 120

Qy  1305 TCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTGGGCCAGTGGTGGCAG 1364
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCATGGACCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTGGGCCAGTGGTGGCAG 180

Qy  1365 CTTTCCCTGGCCTGCGCCGGTGCCACATGCCGTCTCCCACAGTGTGGGCCGTGGTGCAGCTT 1424
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTTTCCCTGGCCTGCGCCGGTGCCACATGCCGTCTCCCACAGTGTGGGCCGTGGTGCAGCTT 240

Qy  1425 CAGCCGCCCCTCACCGGGAGAAGCAGGTTCCTGCGCCAAATACCGAGGGGACACTGGCCT 1484
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAGCCGCCCCTCACCGGGAGAAGCAGGTTCCTGCGCCAAATACCGAGGGGACACTGGCCT 300

Qy  1485 CCCTCTACCACCGGGACAGCAGCCTGATGACCAGCTTCCTGCCAGCCCTAAGCCTGGAGGTG 1544
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCCTCTACCACCGGGACAGCAGCCTGATGACCAGCTTCCTGCCAGCCCTAAGCCTGGAGGTG 360

Qy  1545 CTAGCAGTGAGGACAGCAGCCTGATGACCAGCAGCTTCCTGAGCTCCCTGAGCTCCCT 1604
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  361  CTAGCAGTGAGGAGCAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTCCACCCGAGCTCCCT  420

Qy 1605  TCCCTAATGACACGTGGGTGCTGGAGGCAGTGCCTGCTCCTCCACCCGCGCTCT 1664
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  TCCCTAATGACACGTGGGTGCTGGAGGCAGTGCCTGCTCCTCCACCCGCGCTCT  480

Qy 1665  GCGGGGCCTCGCCTGTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCGAGGCCA 1724
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GCGGGGCCTCGCCTGTGATGTCTCCGTACGTGTGGTGGGTGAGCCACCGAGGCCA  540

Qy 1725  GGTGGTTCCGGGCCGGGGCATCGGCATCCGGACCTCGCCATCCTGGATAGTGCCTTCCTGC 1784
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  GGGTGGTTCCGGGCCGGGGCATCCGGACCTCGCCATCCTGGATAGTGCCTTCCTGC  600

Qy 1785  TGTCCCAGTGTGGCCCATCCTGTTTATGGGCTCCATTGCCCAGCTCAGCCAGTCTGTCA 1844
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  TGTCCCAGGTGTGGCCCCATCCGTTTATGGGCTCCATTGCCCAGCTCAGCCAGTCTGTCA  660

Qy 1845  CTGCCTATATGGTGTCTGCGCAGCCGTCGGTCTGGTCTCGGTCTTTGCTACACAGG 1904
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  CTGCCTATATGGTGTCTGCGCAGCCGTCGGTCTGGTCTCGGTCTTTGCTACACAGG  720

Qy 1905  TAGTATTTGACAAGAGGACTTGGCCAAATACTTCAGGTAGAAAACTTCCAGCACATTGG 1964
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  TAGTATTTGACAAGAGGACTTGGCCAAATACTTCAGGTAGAAAACTTCCAGCACATTGG  780

Qy 1965  GGTGGAGGGCCTCCTCACTGGGTCCCCGCTCCTGTTAGCCCCATGGGCTGC 2024
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781  GGTGGAGGGCCTCCTCACTGGGTCCCCGCTCCTGTTAGCCCCATGGGCTGC  840

Qy 2025  CGGGCTGCCGCCCAGTTTCGTTGCTGCCAAAGTAATGTGGCCTCTGCCTGCCACCCGT 2084
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841  CGGGCTGCCGCCCAGTTTCGTTGCTGCCAAAGTAATGTGGCCTCTGCCTGCCACCCGT  900

Qy 2085  GCTGCTGAGGTGCGTAGCTGCACAGCTGGGGGCGTGGGGGCTGGGGGGTCCCCAG 2144
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901  GCTGCTGAGGTGCGTAGCTGCACAGCTGGGGGCGTGGGGGCTGGGGGGTCCCCAG  960

Qy 2145  TCTCTAGGGCTGCTGACTGGAGCCTTCCAAGGGGTTTCAGTCTGACTTATACAGGG 2204
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961  TCTCTAGGGCTGCTGACTGGAGCCTTCCAAGGGGTTTCAGTCTGACTTATACAGGG 1020

Qy 2205  AGGCCAGAGAGGGCTCCCATGCACTCGGGAATGCGGGGACTCGCAGGTGGATTACCCAGGCTC 2264
         ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db 1021  AGGCCAGAGAGGGCTCCCATGCACTCGGGAATGC-GGGACTCTGCAGGTGGATTACCCAGGCTC 1079

Qy 2265  AGGGTTAACAGCTAGCCTCCAGTTGAGACACACTAGAGAAGGGTTTTTGGGAGCTGAA 2324
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1080  AGGGTTAACAGCTAGCCTCCAGTTGAGACACACTAGAGAAGGGTTTTTGGGAGCTGAA 1139

Qy 2325  TAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAAGT 2384
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1140  TAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAAGT 1199

Qy 2385  AGCTCTTGCATGGGGAGTTTCTAGGAGATGAAAACACTCCTCCATGGGATTTGAACATATG--A 2442
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1200  AGCTCTTGCATGGGGAGTTTCTAGGAGATGAAAACACTCCTCCATGGGATTTGAACATATGAAA 1259

Qy 2443  CTTATTTGTAGGGGAAGAGTCCTGAGGGCAACACACCAGGTCCCCTCAGCCCA 2502
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1260  GTTATTTGTAGGGGAAGAGTCCTGAGGGCAACACACCAGGTCCCCTCAGCCCA 1319

Qy 2503  CAGCACTGTCTTTTGCTGATCCACCCCCCTTACCTTTATCAGGATGTGGCCTGTTG 2562
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1320  CAGCACTGTCTTTTGCTGATCCACCCCCCTTACCTTTATCAGGATGTGGCCGTGTTG 1379

Qy 2563  GTCCTTCGTTGCCATCACAGAGACACAGGCATTAAACATATTAATTATTTAACA 2622
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1380  GTCCTTCGTTGCCATCACAGAGACACAGGCATTAAACATATTAATTATTTAACA 1439

Qy 2623  AAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATATTTGGGTAGGGTG 2682
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1440  AAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATATTTGGGTAGGGTG 1499
```

```
Qy 2683  GGGGATCCCAACAATCAGGTCCCCTGAGATAGTGGTCATTGGGCTGATCATTGCCAGA 2742
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1500  GGGGATCCCAACAATCAGGTCCCCTGAGATAGTGGTCATTGGGCTGATCATTGCCAGA 1559

Qy 2743  ATCTTCTTCTTCCTGGGTCTGGCCCCCAAAATGCTAACCCAGGACCTTGGAAATTCTA 2802
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1560  ATCTTCTTCTTCCTGGGTCTGGCCCCCAAAATGCTAACCCAGGACCTTGGAAATTCTA 1619

Qy 2803  CTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAAGGAAGGTA 2862
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1620  CTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAAGGAAGGTA 1679

Qy 2863  GAGGGTGGGGCTTCAGGTCTCCAGGCTCCCTAACCACCCCTCTTCTCTGGCCCAGCC 2922
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1680  GAGGGTGGGGCTTCAGGTCTCCAAGGCTCCCTAACCACCCCTCTTCTCTGGCCCAGCC 1739

Qy 2923  TGGTTCCCCCACTTCCACTCCCCTCTACTTCCTTAGGACTGGGCTGATGAAGGCACTG 2982
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1740  TGGTTCCCCCACTTCCACTCCCCTCTACTTCCTTAGGACTGGGCTGATGAAGGCACTG 1799

Qy 2983  CCCAAAATTCCCCTACCCCAACTTCCCCTACCCCCCAACTTCCCCACCAGCTCCACA 3042
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1800  CCCAAAATTCCCCTACCCCAACTTCCCCTACCCCCCAACTTCCCCACCAGCTCCACA 1859

Qy 3043  ACCCTGTTTGGAGCTACTGAGGACCAGAGCACAAAGTGCGGTTCCCAAGCCTTTGTC 3102
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1860  ACCCTGTTTGGAGCTACTGAGGACCAGAGCACAAAGTGCGGTTCCCAAGCCTTTGTC 1919

Qy 3103  CATCTTCAGCCCCCAGAGTATATCTGTGCTGGGAATCTCACACAGAAACTCAGGAGCAC 3162
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1920  CATCTTCAGCCCCCAGAGTATATCTGTGCTGGGAATCTCACACAGAAACTCAGGAGCAC 1979

Qy 3163  CCCCTGCCTGCAGCTCTAAGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAAT 3222
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1980  CCCCTGCCTGCAGCTCTAAGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAAT 2039

Qy 3223  AATGTCGTCTTATTTATTTAGCGGGTGAATATTTATACTGTAAGTGAGCAATCAGAGT 3282
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2040  AATGTCGTCTTATTTTTTTAGCGGGTGAATATTTATACTGTAAGTGAGCAATCAGAGT 2099

Qy 3283  ATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
         |||||||||||||||||||||||||||||||||||||||||||||
Db 2100  ATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 2143
```

```
RESULT   5
V58487
ID   V58487 standard; cDNA; 789 BP.
AC   V58487;
DT   08-DEC-1998  (first entry)
DE   3' fragment of prostate tumour specific gene L1-12.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03690.
PR   09-FEB-1998;  US-904809.
PR   25-FEB-1997;  US-805596.
PR   01-AUG-1997;  US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 39; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
```

```
Query Match              19.7%;   Score 673.4;  DB 1;  Length 789;
Best Local Similarity    94.2%;   Pred. No. 1.2e-98;
Matches   745;  Conservative    0;  Mismatches   40;  Indels   6;  Gaps   5;

Qy   1341  CAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGCCGGTGCCACATGCCTGTCCC  1400
                 ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CAGTCTATNTGGCCAGTGTGGCAGCTTCCCTGTGCCGGTGCCACATGCCTGTCCC    60

Qy   1401  ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCTTCTCAGCCCTGC  1460
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCTTCTCAGCCCTGC   120

Qy   1461  AGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCA  1520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCA   180

Qy   1521  AATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGC  1580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  AATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGC   240

Qy   1581  CAGGGCCCTAAGCCTCCACCCGGCTCTGCGAGTCTCCGTACGTGTGGTGCTGGAGGCAGTGGCC  1640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CAGGGCCCTAAGCCTCCACCCGGCTCTGCGAGTCTCCGTACGTGTGGTGCTGGAGGCAGTGGCC   300

Qy   1641  TGCTCCCACCTCCACCCGGCCTCTGCGTGATGCTGTGATGCTGTCCCTGTGCTGGGTCTGTGG  1700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  TGCTCCCACCTCCACCCGGCCTCTGCGTGATGCTGTGATGCTGTCCCTGTGCTGGGTCTGTGG   360

Qy   1701  TGGTGGGTGAGCCCACCGAGGCCAGGGTGTTCCGGGCCGCATCGCCTGGACCTCG  1760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TGGTGGGTGAGCCCACCGAGGCCAGGGTGTTCCGGGCCGCATCGCCTGGACCTCG   420

Qy   1761  TGCTCCCACCTTGCGCCGCAGGCCTGTTTATGGGCTCCA  1820
           ||||||||||||||||||||||||||||||||||||||
Db    421  CCATCCTGTCCTGTCGTCCCANGTGGCCCCATCCCTGTTTATGGGCTCCA   479

Qy   1821  TTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGCTCGGGACCTCG  1880
           ||||||||||||||||| |||||||||||||||||||||||||||||||
Db    480  TTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGCTCGGGACCTCG   539

Qy   1881  TCGCCATTTACTTTGCTACACAGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAG  1940
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540  TC-CCATTTACTTTGCTACACAGTANTATTTGACAAGAACGANTTGGCCAAATACTCAG   598

Qy   1941  CGTAGAAAACTTCCAGCA--CATTGGGGTGGAGGGCCTGCCTCACTGGGTCCCAGCTCCC  1998
           ||| ||| |||||||| |  ||||||| ||| |||||||||||||||||| |||||||||
Db    599  CGTTAAAAAATTCCAGCAACATTGGGGGTGGAAGGCCTGCCTCACTGGGT-CCAACTCCC   657

Qy   1999  CGCTCCTGTTAGCCCCATGGGGCTGCCGGGCTGCCGCCGCCAGTTTCGTTGCTGCCCAAAGT  2058
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Query Match              19.7%;   Score 673.4;  DB 1;  Length 789;
Best Local Similarity    94.2%;   Pred. No. 1.2e-98;
Matches   745;  Conservative    0;  Mismatches   40;  Indels   6;  Gaps   5;

Qy   1341  CAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGCCGGTGCCACATGCCTGTCCC  1400
Db      1  CAGTCTATNTGGCCAGTGTGGCAGCTTCCCTGTGCCGGTGCCACATGCCTGTCCC    60

Qy   1401  ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCTTCTCAGCCCTGC  1460
Db     61  ACAGTGTGGCCGTGTGACAGCTTCAGCGCCCTCACCTTCTCAGCCCTGC   120

Qy   1461  AGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCA  1520
Db    121  AGATCCTGCCCTACACACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCA   180

Qy   1521  AATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGC  1580
Db    181  AATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGC   240

Qy   1581  CAGGGCCCTAAGCCTCCGGAGCTCCCTGCGGGGCCTCTGCGGGGCCTCTGGGTGCTGGGCC  1640
Db    241  CAGGGCCCTAAGCCTCCGGAGCTCCCTGCGGGGCCTCGCGGGGCCTCTGGGTGCTGGGCC   300

Qy   1641  TGCTCCCACCTCCACCCGGCCTCTGCGGGGCCTCTGCGTGATGGTCTCCGTACGTGTGG  1700
Db    301  TGCTCCCACCTCCACCCGGCCTCTGCGGGGCCTCTGCGTGATGGTCTCCGTACGTGTGG   360

Qy   1701  TGGTGGGTGAGCCCACCGAGGCCAGGGTGTTCCGGGCCGCATCGCCTGGACCTCG  1760
Db    361  TGGTGGGTGAGCCCACCGANGCCAGGGTGTTCCGGGCCGCATCGCCTGGACCTCG   420

Qy   1761  CCATCCTGGATAGTGCCTTCCTGTCGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCA  1820
Db    421  CCATCCTGGATAGTG-CTTCCTGTCGTCCCANGTGGCCCCATCCCTGTTTATGGGCTCCA   479

Qy   1821  TTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGCTCGGGTCTCGG  1880
Db    480  TTGTCCAGCTCAGCCAGTCTGTCACTGCCTATATGGTGTCTGCCGCAGCTCGGGTCTCGG   539

Qy   1881  TCGCCATTTACTTTGCTACACAGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAG  1940
Db    540  TC-CCATTTACTTTGCTACACAGTANTATTTGACAAGAACGANTTGGCCAAATACTCAG   598

Qy   1941  CGTAGAAAACTTCCAGCA--CATTGGGGTGGAGGGCCTGCCTCACTGGGTCTGG  1998
Db    599  CGTTAAAAAATTCCAGCAACATTGGGGGTGGAAGGCCTGCCTCACTGGGTCTGG   658

Qy   1999  CGCTCCTGTTAGCCCCATGGGGCTGCCGCCGCCGGCCTGCCGCCGCCAGTTTCGTTGCTGCCCAAAGT  2058
Db    658  CGCTCCTGTTAACCCCATGGGGCTGCTGCCACCCTGTTGCTGCCCAAANT   717

Qy   2059  AATGTGGCTCTCTGCTGCCACCCTGTGCT-GCTGAGGTGCGTAGCTGCACAGCTGGGGGC  2117
Db    718  NATGTGGCTCTCTGCTGCCACCCTGTTGCTGCGGTGCNTACNGCNCANCTNGGGGGG   777

Qy   2118  TGGGGCGTCCC  2128
Db    778  TNGGGNGTTCC   788
```

```
Db    658  CGCTCCTCGTTAACCCATGGGCGTGCCGGCTTGGCCGCCAATTTCTGTTGCTGCCAAANT  717
                |||||||||||||||||||||||  ||| |||| |||| ||| |   ||||||   ||||
Qy   2059  AATGTGGCTCTCTGCTGCTGCCACCCTGTGCT-GCTGAGGTGCGTAGCTGCACAGCTGGGGGC  2117
                |||| ||||||||||||||||||||||  |||  ||| || ||||||
Db    718  NATGTGGCTCTCTGCTGCCACCTGTTGCTGGCTGAAGTGCNTACNGCNCANCTNGGGGGG  777

Qy   2118  TGGGGCGTCCC  2128
                | ||| || ||
Db    778  TNGGGNGTTCC  788


RESULT  7
V58488/c
ID    V58488 standard; cDNA; 772 BP.
AC    V58488;
DT    08-DEC-1998  (first entry)
DE    5' fragment of prostate tumour specific gene L1-12.
KW    Prostate tumour specific gene; human; prostate cancer; detection;
KW    therapy; ss.
OS    Homo sapiens.
PN    WO9837418-A2.
PD    27-AUG-1998.
PF    25-FEB-1998; U03690.
PR    09-FEB-1998; US-904809.
PR    25-FEB-1997; US-806596.
PR    01-AUG-1997; US-904809.
PA    (CORI-) CORIXA CORP.
PI    Dillon DC, Xu J;
DR    WPI; 98-480805/41.
PT    Novel human prostate specific tumour protein and fragments - useful
PT    for detecting and treating prostate cancers
PS    Claim 1; Page 39-40; 141pp; English.
CC    This sequence represents a human prostate tumour specific gene, and can
CC    be used in the method of the invention. The method is for detecting
CC    prostate cancer comprises contacting a biological sample with an agent
CC    able to bind an immunogenic portion of a prostate protein (such as
CC    encoded by this sequence). An antibody which binds to an immunogenic
CC    portion of the prostate protein, and the method can be used to detect,
CC    monitor progression of, or treat prostate cancers. The antibody may
CC    also be conjugated to a therapeutic agent for use in therapy of prostate
CC    cancers.
SQ    Sequence 772 BP;     221 A;     202 C;     170 G;     162 T;

Query Match           17.7%;  Score 603; DB 1; Length 772;
Best Local Similarity  91.8%;  Pred. No. 1.6e-87;
Matches  675;  Conservative    0; Mismatches  49; Indels  11; Gaps  4;

Qy   1961  TTGGGGTGGAGGGCCTGCCTCACTGGGTCCCAGCTCCCGGTCCT-----GTTAGCCCCA  2015
                | |||||| |||||||||||||| | ||  || | | |||    | ||||
Db    735  TNGGGGTGGNGGGGCCTGCCTNAATTGGGTTCCCANGTTNCCCNGNTTCCTGTTAACCCCCN  676

Qy   2016  TGGGGCTGCCGGGCTTGGCCGCC---AGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCT-  2072
                | ||| |||||||| |||||||      |||| ||| || ||| |||||| |||| | | |
Db    675  TNGGGGTGCCGGGGTTGCCCCCCCNAGTTTGTTGTTGTGCCAAAGTAAAGTGGTTTTTTT  616

Qy   2072  GCTGCCACCCTGTCGTCGTGAGGTGCGTAGCTGCCACAGCTGGGGCGTCCCTCT  2131
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    615  GCTGCCACCCTGTCGTCGTGAGGTGCGTAGCTGCACAGCTGGGGCGTCCCTCT  556

Qy   2132  CCCTCTCCCCAGCTCTCAGGGTTCTAGGGCCTGCACTGGAGGCCTTCCAAGGGGGTTTCAGTCTG  2191
                | | |||| |||||||| | |||||||||||||||||||||||||||||||||||||||||
Db    555  CTTTTTTCCCCAGTTTNTAGGGCCTGCACTGGAGNCCTTCCAAGGGGGTTTCAGTCTG  496

Qy   2192  GACTTATACAGGGAGGCCAGAAGGGCTCCATGCACTGGAATGCGGGACTCTGCAGGTGG  2251
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495  GACTTATACAGGGAGGCCAGAAGGGTTCCATGCACTGGAATGCGGGACTCTGCAGGTGG  436

Qy   2252  ATTACCCAGGCTCAGGGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTT  2311
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435  ATTACCCAGGCTCAGGGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTT  376

Qy   2312  TTTGGAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAG  2371
```

```
Db    375  TTTGGGAGCTGAATAAACTCAGTCACCTGGTTTCCCATCNTAAGCCCCTTAACCTGCAG  316
                ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Qy   2372  CTTCGTTTAATGTAGCTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCATGGGGATT  2431
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315  CTTCGTTTAATGTAGCTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCATGGGATT  256

Qy   2432  TGAACATATG--ACTTATTTGTAGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAGG  2489
                |||||||||  |  ||||| | |||||||||||||||||||||||||||||||||||||
Db    255  TGAACATATGAAAGTTATTTGTAGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAGG  196

Qy   2490  TCCCCTCAGCCCACCACTGTCTTTTGCTGCTGATCCACCCCCTCTTACCTTTTATCAGG  2549
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195  TCCCCTCAGCCCACCACTGTCTTTTGCTGCTGATCCACCCCCTCTTACCTTTTATCAGG  136

Qy   2550  ATGTGGCCTGTTGGTCCTTCCTGTTGCCATCACAGAGAACACAGGCATTTAAATATTTAACT  2609
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135  ATGTGGCCTGTTGTTGGTCCTTCCTGTTGCCATCACAGAGAACACAGGCATTTAAATATTTAACT  76

Qy   2610  TATTTATTTTAACAAAGTAGAAGGGAATCCATTGCTTTTCTGTTGGTGTCTAATA  2669
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75  TATTTATTTTAACAAAGTAGAAGGGAATCCATTGCTTTTCTGTTGTGTCTAATA  16

Qy   2670  TTTGGGTAGGGTGGG  2684
                |||||||||||||||
Db    15  TTTGGGTAGGGTGGG  1


RESULT  8
V61145/c
ID    V61145 standard; cDNA; 772 BP.
AC    V61145;
DT    06-JAN-1999  (first entry)
DE    5' cDNA sequence of prostate tumour clone L1-12.
KW    Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS    Homo sapiens.
PN    WO9837093-A2.
PD    27-AUG-1998.
PF    25-FEB-1998; U03492.
PR    09-FEB-1998; US-020956.
PR    25-FEB-1997; US-806099.
PR    01-AUG-1997; US-904804.
PA    (CORI-) CORIXA CORP.
PI    Dillon DC, Xu J;
DR    WPI; 98-609886/51.
PT    Polypeptides comprising immunogenic portions of prostate proteins -
PT    used in a vaccine for the treatment of prostate cancer
PS    Claim 3; Page 39; 130pp; English.
CC    The present sequence is a new DNA which encodes an immunogenic portion
CC    of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC    can be used as a vaccine for the treatment of prostate cancer. The DNA
CC    was identified by analysis of a subtracted cDNA library obtained by
CC    subtracting a prostate tumour cDNA expression library with a normal
CC    tissue cDNA library.
SQ    Sequence 772 BP;     221 A;     202 C;     170 G;     162 T;

Query Match           17.7%;  Score 603; DB 1; Length 772;
Best Local Similarity  91.8%;  Pred. No. 1.6e-87;
Matches  675;  Conservative    0; Mismatches  49; Indels  11; Gaps  4;

Qy   1961  TTGGGGTGGAGGGCCTGCCTCACTGGGTCCCAGCTCCCGGTCCT-----GTTAGCCCCA  2015
                | |||||| |||||||||||||| | ||  || | | |||    | ||||
Db    735  TNGGGGTGGNGGGGCCTGCCTNAATTGGGTTCCCANGTTNCCCNGNTTCCTGTTAACCCCCN  676

Qy   2016  TGGGGCTGCCGGGCTTGGCCGCC---AGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCT-  2072
                | ||| |||||||| |||||||      |||| ||| || ||| |||||| |||| | | |
Db    675  TNGGGGTGCCGGGGTTGCCCCCCCNAGTTTGTTGTTGTGCCAAAGTAAAGTGGTTTTTTT  616

Qy   2072  GCTGCCACCCTGTCGTCGTGAGGTGCGTAGCTGCCACAGCTGGGGCGTCCCTCT  2131
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    615  GCTGCCACCCTGTCGTCGTGAGGTGCGTAGCTGCACAGCTGGGGNTGGGGGCGTCCCTTTT  556
```

```
CC       differentiation activity, haematopoiesis regulating activity, tissue
CC       growth regulating activity, reproductive hormone regulating activity,
CC       chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC       receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC       activity or other activities. The products can be used in forensic, gene
CC       therapy and chromosome mapping procedures. The sequences can also be used
CC       for obtaining corresponding promoter sequences. The nucleic acids
CC       encoding the signal peptides can be used for directing extracellular
CC       secretion of a polypeptide or the insertion of a polypeptide into a
CC       membrane, or importing a polypeptide into a cell.
SQ       Sequence 435 BP;  126 C;  153 G;  81 A;  69 T;
```

```
Query Match             11.9%;  Score 405.8;  DB 1;  Length 435;
Best Local Similarity   98.3%;  Pred. No. 2.3e-56;
Matches 414;  Conservative  6;  Mismatches  0;  Indels  1;  Gaps  1;

Qy        2 GGAACCAGCCTGCACGCGCTGGCTCCGGGTGACAGCGCCGCCGCCTCGGCCAGGATCTGAG   61
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       16 GGAACCAGCCTGCACGCGCTGGCTCCGGGTGACAGCGCCGCCGCCTCGGCCAGGATCTGAG   75

Qy       62 TGATGAGACGTGTCCCCACTGAGGTGCCCACAGCAGCAGCAGGTGTTGAGCAGTGGGCGAGA  121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       76 TGATGAGACGTGTCCCCACTGAGGTGCCCACAGCAGCAGCAGGTGTTGAGCAGTGGGCTGAGA  135

Qy      122 AGCTGGACCGGCACCAAAGGGCTGGCAGAAATGGGCGGCCTGGCTGATTCCTAGGCAGTTG  181
            |||||||||||||||||||||||||||||:||::|||||||||||||||||||||||||||
Db      136 AGCTGGACCGGCACCAAAGGGCTGGCAGAAATDVGCCGGCCTGGCTGATTCCTAGGCAGTTG  195

Qy      182 GCGGCAGCAAGGAGGAGGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGACGAAGCAGTTCTGG  241
            |:||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      196 GCRGCAGCAAGGAGGAGGAGGCGCGCAGCTTCTGGAGCAGAGCCGAGACGAAGCAGTTCTGG  255

Qy      242 AGTGCCTGAACGGCCCCTGAGCCTGCACCCGCTGGCCCACCATGGTCCAGAGGCGTGG  301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      256 AGTGCCTGAACGGCCCCTGAGCCTGCACCCGCTGGCCCACCATGGTCCAGAGGCGTGG  315

Qy      302 GTGAGCCGCCTGCTCGGCACCGGAAAGCCACCAGCTCTTTCGTGGTCAACCTGCTAACCTTT  361
            |||||||||||||||||||||||||||||||||||||::|||||:||||||||||||||||
Db      316 GTGAGCCGCCTGCTCGGCACCGGAAAGCCAGCTCKKGCTGGKCAACCTGCTAACCTTT  375

Qy      362 GGCCTGGAGGTGTGTTTGGCCGCAGGCATCACCTATGTGCCGCCTTCGCTCGCTGGAAGTG  421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      376 GGCCTGGAGGTGTGTTTGGCCGCAGG-ATCACCTATGTGCCGCCTTCGCTCGCTGGAAGTG  434

Qy      422 G  422
            |
Db      435 G  435
```

```
RESULT 10
V71173
ID   V71173 standard; cDNA; 342 BP.
AC   V71173;
DT   12-FEB-1999  (first entry)
DE   PS108 gene-specific clone 3520833.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9950567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen MA, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
```

```
Qy     2132 CCTCCTCCCCAGTCTCTAGGCTGCCTGACTGAGGCCTTCCAAGGGGTTTCAGTCTG  2191
            || ||||||||| ||||| |||||||  |||||||| |||||||||||||||||||||
Db      555 CTTTTTCCCAGTTNTAGGGCTGCCTGACTGAGNCCTTCCAAGGGGTTTCAGTCTG  496

Qy     2192 GACTTATACAGGGAGGCCAGAAGGGCTCCATGCGACTGGAATGCGGGGACTCTGCAGGTGG  2251
            ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db      495 GACTTATACAGGGAGGCCAGAAGGGTTCCATGCACTGGAATGCGGGGACTCTGCAGGTGG  436

Qy     2252 ATTACCCAGGCTCAGGTTAACAGTCTAGCCTCCTAGTTGAGACACACACTTAGAGAAGGGTT  2311
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      435 ATTACCCAGGCTCAGGTTAACAGTCTAGCCTCCTAGTTGAGACACACACTTAGAGAAGGGTT  376

Qy     2312 TTTGGGAGCTGAATAAACTCAGTCACCTGGTTCCCATCTCTAAGCCCCTTAACCTGCAG  2371
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      375 TTTGGGAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTNTAAGCCCCTTAACCTGCAG  316

Qy     2372 CTTCGTTTAAATGTAGCTCTTGCATGGGAGTTTCTAGGATGAAAACTCCTCCATGGGATT  2431
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      315 CTTCGTTTAAATGTAGCTCTTGCATGGGAGTTTCTAGGATGAAAACTCCTCCATGGGATT  256

Qy     2432 TGAACATATG--ACTTATTTGTAGGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAGG  2489
            |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db      255 TGAACATATGAAGTTATTTGTAGGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAGG  196

Qy     2490 TCCCCTCAGCCCACAGCACTGTCTTTTTGCTGATCCACCCCCCTTTATTATCAGG  2549
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      195 TCCCCTCAGCCCACAGCACTGTCTTTTTGCTGATCCACCCCCCTTTATTATCAGG  136

Qy     2550 ATGTGGCCGTTGGTCCTTCTGTTGCCATCACAGAGACACAGGCATTTAAATATTTAACT  2609
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      135 ATGTGGCCGTTGGTCCTTCTGTTGCCATCACAGAGACACAGGCATTTAAATATTTAACT  76

Qy     2610 TATTTATTTAACAAAGTAGAGAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATA  2669
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       75 TATTTATTTAACAAAGTAGAGAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATA  16

Qy     2670 TTTGGGTAGGGTGGG  2684
            |||||||||||||||
Db       15 TTTGGGTAGGGTGGG  1
```

```
RESULT 9
X40506
ID   X40506 standard; cDNA; 435 BP.
AC   X40506;
DT   18-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID NO: 106.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9906550-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1232.
PR   01-AUG-1997; US-905144.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153780/13.
P-PSDB: Y11784.
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
PS   Claim 1; Page 224; 675pp; English.
CC   X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins expressed in prostate, and encode the proteins given in
CC   Y11716 to Y11993 respectively. The proteins given represent the signal
CC   peptide and an N-terminal fragment of a secreted protein. The nucleic
CC   acid sequences can be used for producing secreted human gene products.
CC   They can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell proliferation and
```

```
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence  342 BP;   84 A;   88 C;   81 G;   88 T;

Query Match              9.4%;  Score 319.8;  DB 1;  Length 342;
Best Local Similarity   98.5%;  Pred. No. 8.8e-43;
Matches  333;     Conservative   0;   Mismatches   3;   Indels   2;   Gaps   1;

Qy  2207 GCCAGAAGGGCTCCATGCACTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAG 2266
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GCCAGAAGGGCTCCATGCACTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAG 60

Qy  2267 GGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAAGAGGGTTTTTGGGAGCTGAATA 2326
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAAGAGGGTTTTTGGGAGCTGAATA 120

Qy  2327 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAG 2386
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAG 180

Qy  2387 CTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCATGGGATTTGAACATATG--ACT 2444
         |||||||||||||||||||||||||||||||||||||||||||||| ||||| |    |
Db   181 CTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCATGGGATTTGAACATATGAAAGT 240

Qy  2445 TATTGTAGGGGAAGAGTCCTGAGGGGCAACACAACAGAACCAGGTCCCCTCAGCCCACA 2504
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TATTGTAGGGGAAGAGTCCTGAGGGGCAACACAACAGAACCAGGTCCCCTCAGCCCACA 300

Qy  2505 GCACTGTCTTTTGCTGATCCACCCCCCTTTACCTTT 2542
         ||||||||||||| |||||||||| |||||||||||
Db   301 GCACTGTCTTTTGCTGATCCANCCCCCTCTTACTTTT 338


RESULT   11
V71177
ID   V71177 standard; cDNA; 294 BP.
AC   V71177;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 3705332.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence  294 BP;   70 A;   87 C;   60 G;   74 T;

Query Match              8.5%;  Score 291;  DB 1;  Length 294;
Best Local Similarity   99.0%;  Pred. No. 3.1e-38;
Matches  291;     Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy  2974 AAGGCACTGCCCAAAATTCCCTACCCCCAACTTCCCTACCCCCAACTTCCCCACC 3033
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     1 AAGGCACTGCCCAAAATNNCCCTACCCCCAACTTCCCTACCCCCAACTTCCCCACC 60

Qy  3034 AGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGGAGCACAAAGTGCGGTTTCCCAA 3093
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGGAGCACAAAGTGCGGTTTCCCAA 120

Qy  3094 GCCTTTGTCCATCTCAGCCCCCAGTATATCTGTGCTTGGGAATCTCACACAGAAACT 3153
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCCTTTGTCCATCTCAGCCCCCAGTATATCTGTGCTTGGGAATCTCACACAGAAACT 180

Qy  3154 CAGGAGCACCCCCTGCCTGAGCTAAGGAGGGTCTTATCTCCAGGGGGGTTTAAGTGCC 3213
         ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   181 CAGGAGCACCCCCTGCNTGAGCTAAGGAGGGTCTTATCTCCAGGGGGGTTTAAGTGCC 240

Qy  3214 GTTTGCAATAAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAA 3267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GTTTGCAATAAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAA 294


RESULT   12
V71175
ID   V71175 standard; cDNA; 288 BP.
AC   V71175;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 3497504.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence  288 BP;   70 A;   66 C;   61 G;   90 T;
```

Query Match   8.4%; Score 287; DB 1; Length 288;
Best Local Similarity   99.7%; Pred. No. 1.3e-37;
Matches 287; Conservative   0; Mismatches   1; Indels   0; Gaps   0:

```
Qy 2532 CTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCTTCTGTTGCCATCACAGAGACACAG 2591
        |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    1 CTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCTTCTGTTGCCATCACAGAGACACAG 60

Qy 2592 GCATTTAAATATTTAACTTATTATTTATTAACAAAGTAGAAGGGAATCCATTGCTAGCTTTT 2651
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCATTTAAATATTTAACTTATTATTTATTAACAAAGTAGAAGGGAATCCATTGCTAGCTTTT 120

Qy 2652 CTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGGATCCCAACAATCAGGTCCCCTGAG 2711
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTGTGTTGGTGTCTAATATTTGGGTANGGTGGGGGATCCCAACAATCAGGTCCCCTGAG 180

Qy 2712 ATAGCTCGTCATTGGGCTGATCATTGCCAGAATCTTCTCTCCTGGGGTCTGGCCCCCCA 2771
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ATAGCTCGTCATTGGGCTGATCATTGCCAGAATCTTCTCTCCTGGGGGTCTGGCCCCCCA 240

Qy 2772 AAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAATGATAA 2819
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAATGATAA 288
```

RESULT 13
V71176
ID   V71176 standard; cDNA; 272 BP.
AC   V71176;
DT   12-FEB-1999   (first entry)
DE   PS108 gene-specific clone 3964174.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 272 BP; 58 A; 96 C; 49 G; 68 T;

Query Match   7.9%; Score 270; DB 1; Length 272;
Best Local Similarity   99.6%; Pred. No. 6.4e-35;
Matches 270; Conservative   0; Mismatches   1; Indels   1; Gaps   0:

```
Qy 2795 AAATTCTACTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAA 2854
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AAATTCTACTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAA 60
```

```
Qy 2855 GGAAGGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCTAACCACCCTCTTCCTCCTTG 2914
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGAAGGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCTAACCACCCTCTTCCTCCTTG 120

Qy 2915 GCCCAGCCTGGTTCCCCCCACTTCCACCTTCCACTCCTCTAGGACTGGCGGCCTGATGA 2974
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GCCCAGCCTGGTTCCCCCCACTTCCACCTTCCACTCCTCTAGGACTGGCGGCCTGATGA 180

Qy 2975 AGGCACTGCCCAAATTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCACCACCA 3034
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGGCACTGCCCAAATTCCCTACCCCAACTTTCCCTACCCCAACTTTCCCACCACCA 240

Qy 3035 GCTCCACAACCCTGTTTGGAGCTACTGCAGG 3065
        |||||||||||| |||||||||||||||||
Db  241 GTCCCACAACCCTGTTTGGAGCTACTGCAGG 271
```

RESULT 14
V71174
ID   V71174 standard; cDNA; 265 BP.
AC   V71174;
DT   12-FEB-1999   (first entry)
DE   PS108 gene-specific clone 2188949.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 265 BP; 64 A; 64 C; 61 G; 76 T;

Query Match   7.8%; Score 265; DB 1; Length 265;
Best Local Similarity   100.0%; Pred. No. 4e-34;
Matches 265; Conservative   0; Mismatches   0; Indels   0; Gaps   0:

```
Qy 2453 GGGGAAGAGTCCTGAGGGCAACACACAAGAACCAGGTCCCCTCAGCCCACCAGCACTGTC 2512
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGGGAAGAGTCCTGAGGGCAACACACAAGAACCAGGTCCCCTCAGCCCACCAGCACTGTC 60

Qy 2513 TTTTTGCTGATCCACCCCCTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCTTCTCGT 2572
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTTTTGCTGATCCACCCCCTCTTACCTTTATCAGGATGTGGCCTGTTGGTCCTTCTCGT 120

Qy 2573 TGCCATCACAGAGACACAGGCATTTAATATTTAATATTTAACAAAGTAGAAGG 2632
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TGCCATCACAGAGACACAGGCATTTAATATTTAATATTTAACAAAGTAGAAGG 180
```

Search completed: September 28, 1999, 12:26:53
Job time: 5233 sec

Qy   2633 GAATCCATTGCTAGCTTTTCTGTGTGTCTAATATTGGGTAGGTGGGGGATCCCC 2692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAATCCATTGCTAGCTTTTCTGTGTGTCTAATATTGGGTAGGTGGGGGATCCCC 240

Qy   2693 AACAATCAGGTCCCCTGAGATAGCT 2717
          |||||||||||||||||||||||||
Db   241 AACAATCAGGTCCCCTGAGATAGCT 265

RESULT 15
ID   V71166 standard; cDNA; 258 BP.
AC   V71166;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 1864683.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences v71166-79 represent overlapping clones comprising
CC   the sequence v71180 and the consensus sequence v71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 258 BP; 35 A; 78 C; 88 G; 57 T;

Query Match   7.5%; Score 256.4; DB 1; Length 258;
Best Local Similarity 99.6%; Pred. No. 9e-33;
Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1177 GGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGAGACACTATGA 1236
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGAGACACTATGA 60

Qy   1237 TGAAGGGGTTCGGATGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTT 1296
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGAAGGGGTTCGGATGGCAGCCTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTT 120

Qy   1297 CTCTCTGGTCATGGACCCGCTGGTGCAGGCAGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 1356
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTCTCTGGTCATGGACCCGCTGGTGCAGGCAGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 180

Qy   1357 TGTGGCAGCTTTCCCTGGCTGCCGGTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGT 1416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGTGGCAGCTTTCCCTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGT 240

Qy   1417 GACAGCTTCAGCCGCCCT 1434
          ||||||||||||||||| |
Db   241 GACAGCTTCAGCCGCCTT 258

us-09-030-606-110.rng

GenCore version 4.5

Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 10:58:40 ; Search time 161.06 Seconds
(without alignments)
1940.998 Million cell updates/sec

Title: US-09-030-606-110
Perfect score: 3410
Sequence: 1 GGGAACCAGCCTGCACGCGC.........AAAAAAATAAAAAAAAAAAA 3410

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	91.4	2.7	1066	3	US-08-157-101A-4	Sequence 4, Appli
2	90.8	2.7	6671	2	US-08-280-443-1	Sequence 1, Appli
3	90.8	2.7	6671	2	US-08-457-459-1	Sequence 1, Appli
4	90.8	2.7	6671	5	US-08-555-678-1	Sequence 1, Appli
5	90.8	2.7	6671	5	PCT-US95-02275-1	Sequence 1, Appli
6	89.4	2.6	144	2	US-08-702-344-26	Sequence 26, Appli
7	89	2.6	1641	1	US-08-300-903A-8	Sequence 8, Appli
8	88.2	2.6	140	1	US-08-628-417-5	Sequence 5, Appli
9	88.2	2.6	240	1	US-08-628-417-6	Sequence 6, Appli
10	88.2	2.6	1602	3	US-08-530-950-3	Sequence 3, Appli
11	87.8	2.6	1493	1	US-08-340-820-24	Sequence 24, Appl
12	87.8	2.6	1493	1	US-08-593-535-24	Sequence 24, Appl
13	86.6	2.5	1817	1	US-08-473-981A-5	Sequence 5, Appli
14	85	2.5	903	5	PCT-US95-06406A-21	Sequence 21, Appli
15	84	2.5	2082	4	US-08-785-310A-2	Sequence 2, Appli
16	83.8	2.5	635	3	US-08-455-633A-35	Sequence 35, Appli
17	83.8	2.5	635	3	US-08-416-336-5	Sequence 5, Appli
18	83.8	2.5	635	5	PCT-US94-05354-35	Sequence 35, Appli
19	83.4	2.4	630	1	US-08-185-414E-1	Sequence 1, Appli
c 20	83	2.4	9589	1	US-07-925-695-1	Sequence 1, Appli
c 21	83	2.4	9589	1	US-07-925-695-2	Sequence 2, Appli
22	83	2.4	98	2	US-08-088-658-42	Sequence 42, Appli
23	83	2.4	117	2	US-08-702-344-3	Sequence 3, Appli
24	81.2	2.4	1332	4	US-09-057-762-1	Sequence 1, Appli
25	81	2.4	1172	1	US-07-945-288-9	Sequence 9, Appli
26	81	2.4	1172	1	US-08-462-831-9	Sequence 9, Appli
27	81	2.4	1172	3	US-08-461-809-9	Sequence 9, Appli
28	81	2.4	1172	3	US-08-461-441-9	Sequence 9, Appli
29	81	2.4	1172	5	PCT-US93-05518-9	Sequence 9, Appli
c 30	79.6	2.3	270	4	US-08-520-678A-30	Sequence 30, Appl
31	79.2	2.3	260	4	US-08-520-678A-29	Sequence 29, Appl
c 32	78.2	2.3	356	4	US-08-520-678A-22	Sequence 22, Appl
33	77.8	2.3	1147	3	US-08-665-716-1	Sequence 1, Appli
34	77.8	2.3	84	3	US-08-664-596B-3	Sequence 3, Appli
35	77.6	2.3	84	3	US-08-738-167-3	Sequence 3, Appli
36	77.6	2.3	1921	4	US-08-557-128-11	Sequence 11, Appli
37	76.8	2.3	11517	3	US-07-920-281C-1	Sequence 1, Appli
38		76.6		2.2	2010	1	US-07-864-475A-4	Sequence 4, Appli
39		76.6		2.2	2010	4	US-08-468-249A-4	Sequence 4, Appli
40		76.4		2.2	1023	2	US-08-252-966B-16	Sequence 16, Appl
41		76.4		2.2	5173	2	US-08-242-677-1	Sequence 1, Appli
42		76.4		2.2	1325	4	US-08-464-517-1	Sequence 1, Appli
43		76.2		2.2	1325	5	PCT-US93-05000-1	Sequence 1, Appli
44		76.2		2.2	5852	1	US-07-867-106-2	Sequence 2, Appl
c 45		76.2		2.2	1325	3	US-08-306-691B-51	Sequence 51, Appl

ALIGNMENTS

RESULT 1
US-08-157-101A-4
; Sequence 4, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; TITLE OF INVENTION: PLASMIDS THEREFOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PILLSBURY, MADISON & SUTRO
;   STREET: 1100 NEW YORK AVENUE, N.W.
;   CITY: WASHINGTON
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/157,101A
;   FILING DATE: 05-APR-1994
;   CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
;   NAME: TITUS, MARIANA K
;   REGISTRATION NUMBER: 35843
;   REFERENCE/DOCKET NUMBER: 9437/204199
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-861-3711
;   TELEFAX: 202-822-0944
;   TELEX: 671627 CUCH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1066 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-157-101A-4

Query Match   2.7%;   Score 91.4; DB 3; Length 1066;
Best Local Similarity 81.9%; Pred. No. 7.2e-10;
Matches 104; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

Qy 3284 TAATGTTTATGGTGACAAATTAAAGGCTTTCTTATATGTTAAAAAAAAAAAAAAAA 3343
         |||||| || ||| |||| || | | ||| |||||||||||||||||||||||||
Db 898 TAATGTTGGAGGAGAATGAATAAATAAAGTGAATCTTTGCAAAAAAAAAAAAAAAA 957

Qy 3344 AAAAAAAAAAAAAAAAAAAARARARAAAAAAAAAAAAAAAAAAATAAAA 3403
         ||||||||||||||||||::|||||||||||||||||||||| ||||
Db 958 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1017

Qy 3404 AAAAAAA 3410

Wed Sep 29 14:26:55 1999

```
Db   1018 AAAAAAA 1024
          |||||||
```

RESULT 2
US-08-280-443-1
; Sequence 1. Application US/08280443
; Patent No. 5643778
, GENERAL INFORMATION:
,   APPLICANT: Nishikura, Kazuko
,   TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
,   TITLE OF INVENTION: Thereof
,   NUMBER OF SEQUENCES: 39
,   CORRESPONDENCE ADDRESS:
,   ADDRESSEE: Howson and Howson
,   STREET: Spring House Corporate Cntr, P.O. Box 457
,   CITY: Spring House
,   STATE: Pennsylvania
,   COUNTRY: USA
,   ZIP: 19477
,   COMPUTER READABLE FORM:
,   MEDIUM TYPE: Floppy disk
,   COMPUTER: IBM PC compatible
,   OPERATING SYSTEM: PC-DOS/MS-DOS
,   SOFTWARE: PatentIn Release #1.0, Version #1.25
,   CURRENT APPLICATION DATA:
,   APPLICATION NUMBER: US/08/280,443
,   FILING DATE:
,   CLASSIFICATION: 435
,   PRIOR APPLICATION DATA:
,   APPLICATION NUMBER: US 08/197,794
,   FILING DATE: 17-FEB-1994
,   ATTORNEY/AGENT INFORMATION:
,   NAME: Bak, Mary E.
,   REGISTRATION NUMBER: 31,215
,   REFERENCE/DOCKET NUMBER: WST49AUSA
,   TELECOMMUNICATION INFORMATION:
,   TELEPHONE: 215-540-9206
,   TELEFAX: 215-540-5818
, INFORMATION FOR SEQ ID NO: 1:
,   SEQUENCE CHARACTERISTICS:
,   LENGTH: 6671 base pairs
,   TYPE: nucleic acid
,   STRANDEDNESS: double
,   TOPOLOGY: unknown
,   MOLECULE TYPE: cDNA
,   FEATURE:
,   NAME/KEY: CDS
,   LOCATION: 155..3832
US-08-280-443-1

Query Match   2.7%;   Score 90.8; DB 2; Length 6671;
Best Local Similarity 76.4%; Pred. No. 1.7e-09;
Matches 110;   Conservative   1; Mismatches 33; Indels   0; Gaps   0;

Qy   3267 AGTGAGCAATCAGAGTATAATGTTTATGGTGACAAATTAAGGCTTTCTATATGTTTA 3326
          |||| | | || || || | | || | |||| |
Db   6522 AGTGACTTAACAATATACATTCCTCATAAATAAAAAAAACAAGAATCTGAAAAAAAA 6581

Qy   3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
          ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db   6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6641

Qy   3387 AAAAAAAAAATAAAAAAAAAAA 3410
          |||||||||| |||||||||||
Db   6642 AAAAAAAAAAAAAAAAAAAAAA 6665
```

RESULT  3
US-08-457-459-1
; Sequence 1. Application US/0845745Z
; Patent No. 5677428

```
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Howson and Howson
;           STREET: Spring House Corporate Cntr, P.O. Box 457
;           CITY: Spring House
;           STATE: Pennsylvania
;           COUNTRY: USA
;           ZIP: 19477
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/08/555,678
;           FILING DATE:
;           CLASSIFICATION: 435
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/197,794
;           FILING DATE: 17-FEB-1994
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/280,443
;           FILING DATE: 25-JUL-1994
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/457,459
;           FILING DATE: 01-JUN-1995
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Bak, Mary E.
;           REGISTRATION NUMBER: 31,215
;           REFERENCE/DOCKET NUMBER: WST49DUSA
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 215-540-9206
;           TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 6671 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       FEATURE:
;           NAME/KEY: CDS
;           LOCATION: 155..3832
US-08-555-678-1

Query Match     2.7%;      Score 90.8;  DB 3;  Length 6671;
Best Local Similarity  76.4%;      Pred. No. 1.7e-09;
Matches 110;  Conservative  1;  Mismatches  33;  Indels  0;  Gaps  0;

Qy  3267 AGTGAGCAATCAGAGTATATATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
         |||||  || ||  |||  || |||| ||  |||| ||
Db  6522 AGTGACTTAACATATACATTCCTCATAAATAAAGGCTTTCTTATATGTTTA 6581

Qy  3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
         ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db  6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6641

Qy  3387 AAAAAAAAAAATAAAAAAAAAAA 3410
         ||||||||| |||||||||
Db  6642 AAAAAAAAAAAAAAAAAAAAAAA 6665

RESULT 5
PCT-US95-02275-1
; Sequence 1, Application PC/TUS9502275
; GENERAL INFORMATION:
;       APPLICANT: Wistar Institute of Anatomy & Biology
;       TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
;       TITLE OF INVENTION: Thereof
;       NUMBER OF SEQUENCES: 39
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Howson and Howson
;           STREET: Spring House Corporate Cntr, P.O. Box 457
```

```
;           CITY: Spring House
;           STATE: Pennsylvania
;           COUNTRY: USA
;           ZIP: 19477
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: PCT/US95/02275
;           FILING DATE:
;           CLASSIFICATION:
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/280,443
;           FILING DATE: 25-JUL-1994
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US 08/197,794
;           FILING DATE: 17-FEB-1994
;       ATTORNEY/AGENT INFORMATION:
;           NAME: Bak, Mary E.
;           REGISTRATION NUMBER: 31,215
;           REFERENCE/DOCKET NUMBER: WST49BPCT
;       TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 215-540-9206
;           TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 6671 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: unknown
;       MOLECULE TYPE: cDNA
;       FEATURE:
;           NAME/KEY: CDS
;           LOCATION: 155..3832
PCT-US95-02275-1

Query Match     2.7%;      Score 90.8;  DB 5;  Length 6671;
Best Local Similarity  76.4%;      Pred. No. 1.7e-09;
Matches 110;  Conservative  1;  Mismatches  33;  Indels  0;  Gaps  0;

Qy  3267 AGTGAGCAATCAGAGTATATATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
         |||||  || ||  |||  || |||| ||  |||| ||
Db  6522 AGTGACTTAACATATACATTCCTCATAAATAAAGGCTTTCTTATATGTTTA 6581

Qy  3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
         ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db  6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6641

Qy  3387 AAAAAAAAAAATAAAAAAAAAAA 3410
         ||||||||| |||||||||
Db  6642 AAAAAAAAAAAAAAAAAAAAAAA 6665

RESULT 6
US-08-702-344-26
; Sequence 26, Application US/08702344
; Patent No. 5723315
; GENERAL INFORMATION:
;       APPLICANT: Jacobs, Kenneth
;       APPLICANT: McCoy, John
;       APPLICANT: LaVallie, Edward
;       APPLICANT: Racie, Lisa
;       APPLICANT: Merberg, David
;       APPLICANT: Treacy, Maurice
;       APPLICANT: Spaulding, Vikki
;       TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;       TITLE OF INVENTION: ENCODING THEM
;       NUMBER OF SEQUENCES: 37
;       CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Genetics Institute, Inc.
```

```
      STREET:   87 CambridgePark Drive
      CITY:   Cambridge
      STATE:   Massachusetts
      COUNTRY:  U.S.A.
      ZIP:   02140
   COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/702,344
      FILING DATE:
      CLASSIFICATION:  536
   ATTORNEY/AGENT INFORMATION:
      NAME:  Brown, Scott A.
      REGISTRATION NUMBER:  32,724
   TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (617) 498-8224
      TELEFAX:  (617) 876-5851
INFORMATION FOR SEQ ID NO: 26:
   SEQUENCE CHARACTERISTICS:
      LENGTH:  144 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  double
      TOPOLOGY:  linear
   MOLECULE TYPE:  cDNA
US-08-702-344-26

  Query Match     2.6%;  Score 89.4;  DB 2;  Length 144;
  Best Local Similarity  85.0%;  Pred. No. 9.1e-10;  Indels 0; Gaps 0;
  Matches 96;  Conservative 1;  Mismatches 1;

Qy  3298 ACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAA 3357
          | |||| ||  ||||| ||    ||||||||||||||||||||||||||||
Db     1 AAAAAAAAAANNTNNTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy  3358 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA 3410
          |||||||||||:|||||||||||||||| |||||||||
Db    61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 113

RESULT 7
US-08-300-903A-8
; Sequence 8, Application US/08300903A
; Patent No. 5591630
; GENERAL INFORMATION:
      APPLICANT:  Anderson, Dirk M
      APPLICANT:  Giri, Judith G
      TITLE OF INVENTION:  Interleukin-15 Receptors
      NUMBER OF SEQUENCES:  15
      CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Immunex Corporation
      STREET:  51 University Street
      CITY:  Seattle
      STATE:  Washington
      COUNTRY:  USA
      ZIP:  98101
   COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  Apple Macintosh
      OPERATING SYSTEM:  Apple Operating System 7.1
      SOFTWARE:  Microsoft Word for Apple, Version 5.1a
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/300,903A
      FILING DATE:  06-SEPTEMBER-1994
      CLASSIFICATION:  435
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  USSN 08/236,919
      FILING DATE:  06-MAY-1994
      CLASSIFICATION:  435
   ATTORNEY/AGENT INFORMATION:
```

```
      NAME:  Perkins, Patricia Anne
      REGISTRATION NUMBER:  34,695
      REFERENCE/DOCKET NUMBER:  2822-A
   TELECOMMUNICATION INFORMATION:
      TELEPHONE:  206-587-0430
      TELEFAX:  206-233-0644
INFORMATION FOR SEQ ID NO: 8:
   SEQUENCE CHARACTERISTICS:
      LENGTH:  1641 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
   MOLECULE TYPE:  cDNA
   HYPOTHETICAL:  NO
   ANTI-SENSE:  NO
   FEATURE:
      NAME/KEY:  CDS
      LOCATION:  3..839
US-08-300-903A-8

  Query Match     2.6%;  Score 89;  DB 1;  Length 1641;
  Best Local Similarity  85.2%;  Pred. No. 2.5e-09;  Indels 0; Gaps 0;
  Matches 98;  Conservative 1;  Mismatches 16;

Qy  3296 TGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAA 3355
           |  ||| |||| ||   | | |||   ||||||||||||||||||||||||
Db  1525 TAAACAAAATAAAACATCTATTTTCAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1584

Qy  3356 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA 3410
           ||||||||||||||||||||||||||||||||||||| ||||||||||
Db  1585 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1639

RESULT 8
US-08-628-417-5
; Sequence 5, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
      APPLICANT:  GILLESPIE, DAVID
      TITLE OF INVENTION:  COMPETITOR PRIMER ASYMMETRIC
      TITLE OF INVENTION:  POLYMERASE CHAIN REACTION
      NUMBER OF SEQUENCES:  7
      CORRESPONDENCE ADDRESS:
      ADDRESSEE:  U.S. ARMY CHEMICAL AND BIOLOGICAL
      ADDRESSEE:  DEFENSE COMMAND
      STREET:  OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
      CITY:  ABERDEEN PROVING GROUND
      STATE:  MARYLAND
      COUNTRY:  USA
      ZIP:  21010-5423
   COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  IBM PC compatible
      OPERATING SYSTEM:  PC-DOS/MS-DOS
      SOFTWARE:  PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/628,417
      FILING DATE:
      CLASSIFICATION:  435
   ATTORNEY/AGENT INFORMATION:
      NAME:  BIFFONI, ULYSSES J
      REGISTRATION NUMBER:  39,908
      REFERENCE/DOCKET NUMBER:  DAM 398-94
   TELECOMMUNICATION INFORMATION:
      TELEPHONE:  410-671-1158
      TELEFAX:  410-671-2534
INFORMATION FOR SEQ ID NO: 5:
   SEQUENCE CHARACTERISTICS:
      LENGTH:  140 bases
      TYPE:  nucleic acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
```

; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-628-417-5

Query Match   2.6%;   Score 88.2;   DB 1;   Length 140;
Best Local Similarity   80.3%;   Pred. No. 1.6e-09;
Matches 102;   Conservative 1;   Mismatches 24;   Indels 0;   Gaps 0;

Qy 3284 TAATGTTTATGGTGACAAAATAAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAA 3343
        || | | | |||| | || | | ||| || | || | |||||||||||||||||||||
Db   1 TAGAGATAAAGACAATAAAACTTTAGAAATAATTTTACTAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy 3344 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAA 3403
        ||||||||||||||||||||||||||||:||||||||||||||||||||| |||||
Db   61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy 3404 AAAAAA 3410
        ||||||
Db 121 AAAAAA 127

RESULT 9
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULYSSES J
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-628-417-6

Query Match   2.6%;   Score 88.2;   DB 1;   Length 240:
Best Local Similarity   80.3%;   Pred. No. 1.9e-09;
Matches 102;   Conservative 1;   Mismatches 24;   Indels 0;   Gaps 0;

Qy 3284 TAATGTTTATGGTGACAAAATAAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAA 3343

Db || | | | |||| | | | || || | || | | || ||||||||||||||||||||||
    1 TAGAGATAAAGACAATAAAACTTTAGAAATAATTTTACTAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy 3344 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAA 3403
        ||||||||||||||||||||||||||||::|||||||||||||||||||||||||| ||||||
Db   61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy 3404 AAAAAA 3410
        ||||||
Db 121 AAAAAA 127

RESULT 10
US-08-530-950-3
; Sequence 3, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-530-950-3

Query Match   2.6%;   Score 88.2;   DB 3;   Length 1602;
Best Local Similarity   86.5%;   Pred. No. 3.5e-09;
Matches 96;   Conservative 1;   Mismatches 14;   Indels 0;   Gaps 0;

Qy 3300 AAAATTAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3359
        |||||| || | ||| ||||| ||||||||||||||||||||||||||||||||||||||||
Db 1490 AATATTTAATGATGTGTCATATGAGTCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1549

Qy 3360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAA 3410
        |||||||||||:||||||||||||||||||||||||||||||| |||||||||
Db 1550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1600

RESULT 11
US-08-340-820-24

Wed Sep 29 14:26:55 1999

Sequence 24, Application US/08340820
Patent No. 5512460
GENERAL INFORMATION:
     APPLICANT: NARUO, Ken-ichi
     APPLICANT: SEKO, Chisako
     APPLICANT: KUROKAWA, Tsutomu
     APPLICANT: KONDO, Tatsuya
     TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
     TITLE OF INVENTION: PRODUCTION
     NUMBER OF SEQUENCES: 27
     CORRESPONDENCE ADDRESS:
          ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
          ADDRESSEE: CUSHMAN
          STREET: 130 Water Street
          CITY: Boston
          STATE: Massachusetts
          COUNTRY: US
          ZIP: 02109
     COMPUTER READABLE FORM:
          MEDIUM TYPE: Floppy disk
          COMPUTER: IBM PC compatible
          OPERATING SYSTEM: PC-DOS/MS-DOS
          SOFTWARE: PatentIn Release #1.0, Version #1.25
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/340,820
          FILING DATE:
          CLASSIFICATION: 435
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US/07/835,713
          FILING DATE:
     ATTORNEY/AGENT INFORMATION:
          NAME: CONLIN, David G.
          REGISTRATION NUMBER: 27026
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: (617)523-3400
          TELEFAX: (617)523-6440
          TELEX: 200291 STRE UR
     INFORMATION FOR SEQ ID NO: 24:
          SEQUENCE CHARACTERISTICS:
               LENGTH: 1493 base pairs
               TYPE: nucleic acid
               STRANDEDNESS: double
               TOPOLOGY: linear
          MOLECULE TYPE: cDNA to mRNA
          HYPOTHETICAL: NO
          ANTI-SENSE: NO
          ORIGINAL SOURCE:
               ORGANISM: Homo sapiens
               HAPLOTYPE: 2n
               TISSUE TYPE: skin
               CELL TYPE: fibroblast
          IMMEDIATE SOURCE:
               LIBRARY: Human foreskin cDNA library
               CLONE: pGAF1
US-08-340-820-24

Query Match   2.6%; Score 87.8; DB 1; Length 1493;
Best Local Similarity   75.9%; Pred. No. 4.1e-09;
Matches 107; Conservative   1; Mismatches   33; Indels   0; Gaps   0;

Qy 3264 GTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTCTTATATGT 3323
              ||| || | ||| | | |||| |||||| ||| |||||
Db 1353 GTAAAAATAAAAAAAAATAAAAAAAGTTAAATTTATTTATAGAAAT 1412
Qy 3324 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3383
              | |||||||||||||||||||||||||||||||||||||||||||
Db 1413 TCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1472
Qy 3384 AAAAAAAAAAAAATAAAAA 3404
              ||||||||||| |||||
Db 1473 AAAAAAAAAAAAAAAAAAA 1493

```
Db   1473 AAAAAAAAAAAAAAAAAAA 1493
```

RESULT 13
US-08-473-981A-5
; Sequence 5, Application US/08473981A
; Patent No. 5629162
; GENERAL INFORMATION:
; APPLICANT: deFougerolles, Antonin R
; APPLICANT: Springer, Timothy A
; TITLE OF INVENTION: METHODS OF IDENTIFYING AGENTS WHICH MODULATE
; TITLE OF INVENTION: ICAM-3 BINDING TO LFA-1 (AS AMENDED)
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,981A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLONIG, ROBERT C
; REGISTRATION NUMBER: 34,395
; REFERENCE/DOCKET NUMBER: 1011.0560004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1817 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..1649
US-08-473-981A-5

```
Query Match           2.5%;  Score 86.6;  DB 1;  Length 1817;
Best Local Similarity 85.6%;  Pred. No. 7.6e-09;
Matches  95; Conservative   1; Mismatches  15; Indels   0; Gaps   0;

Qy   3299 CAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3358
          || | |||||| ||    |    |||||||||||||||||||||||||||||||||||||||
Db   1707 CACCAATAAAGCCTCAAACTCCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1766

Qy   3359 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAA 3409
          ||||||||:||||||||||||||||||||||||||||| |||||||||
Db   1767 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1817
```

RESULT 14
PCT-US95-06406A-21
; Sequence 21, Application PCT/US9506406A
; GENERAL INFORMATION:
; APPLICANT: Janet D. Robishaw, Charles Kunsch
; TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
; TITLE OF INVENTION: Subunits
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06406A
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: No
PCT-US95-06406A-21

```
Query Match           2.5%;  Score 85;  DB 5;  Length 903;
Best Local Similarity 84.7%;  Pred. No. 1.2e-08;
Matches  94; Conservative   1; Mismatches  16; Indels   0; Gaps   0;

Qy   3300 AAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3359
          || |:|||||| |    |    ||||||||||||||||||||||||||||||||||||||||
Db    788 AAGAATAAAGTCATCCAGAGCCTCAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAA 847

Qy   3360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAA 3410
          ||||||||:||||||||||||||||||||||||||||||||||| ||||||||||
Db    848 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 898
```

RESULT 15
US-08-785-310A-2
; Sequence 2, Application US/08785310A
; Patent No. 5840532
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Neuronal PAS Domain Protein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,310A
; FILING DATE: .21-JAN-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:

us-09-030-606-110.rni

```
     NAME: OSMAN, RICHARD A
     REGISTRATION NUMBER: 36,627
     REFERENCE/DOCKET NUMBER: UTSD:1226
   TELECOMMUNICATION INFORMATION:
     TELEPHONE: (415) 343-4341
     TELEFAX: (415) 343-4342
 INFORMATION FOR SEQ ID NO: 2:
   SEQUENCE CHARACTERISTICS:
     LENGTH: 2082 base pairs
     TYPE: nucleic acid
     STRANDEDNESS: double
     TOPOLOGY: linear
     MOLECULE TYPE: cDNA
US-08-785-310A-2
```

```
Query Match            2.5%;   Score 84;  DB 4;  Length 2082;
Best Local Similarity  88.2%;  Pred. No. 2.6e-08;
Matches  90;  Conservative   1;  Mismatches   1;  Indels   0;  Gaps   0;

Qy  3309 GGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3368
              ||| ||| |   |||||||||||||||||||||||||||||||||||||||||||||
Db  1953 GCCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2012

Qy  3369 RAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA 3410
         :|||||||||||||||||||||||||| |||||||||||||
Db  2013 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2054
```

Search completed: September 28, 1999, 11:33:02
Job time: 2062 sec

OM of: US-09-030-606-110 to: PIR_60:*   out_format : pfs

Date: Sep 25, 1999   6:25 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (C) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame_n2p.model -DEV-xlp
-Q=/cgn2.1/USPTO_spool/US09030606/runat_24091999_171617_29825/app_query.fasta.1
-DB-PIR_60 -QFMT-fastan -SUFFIX=pr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH-0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:
Query: US-09-030-606-110
Query length: 3410
Database: PIR_60:*
Database sequences: 122810
Database length: 40065486
Search time (sec): 182.540000

pir1:CGHU1D   +   304.50   286.66   4.3e-09   680 ! collagen alpha 1(X) chain pr
pir1:CGHU2B   +   304.00   279.34   4.4e-09   1712 ! collagen alpha 2(IV) chain
pir1:CGHU3B   +   304.00   279.52   4.4e-09   1670 ! collagen alpha 3(IV) chain

seq_name: pir1:QQBE3

seq_documentation_block:
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C;Species: human herpesvirus 4, Epstein-Barr virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C;Accession: A03742
R;Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A;Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi
A;Reference number: A93065; MUID:85035713
A;Accession: A03742
A;Molecule type: DNA
A;Residues: 1-660 <BAN>
R;Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.
Nature 310, 207-211, 1984
A;Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A;Reference number: A03794; MUID:84270667
A;Contents: annotation; protein coding region
C;Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399
C;Superfamily: human herpesvirus 4 BHLF1 protein

alignment_scores:
              Quality: 398.50   Length: 860
              Ratio: 1.126   Gaps: 54
         Percent Similarity: 41.163   Percent Identity: 27.558

alignment_block:
US-09-030-606-110/rev x QQBE3

Align seg 1/1 to: QQBE3 from: 1 'to: 660

2297 TGTGTCTCAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGTAATCCAC 2248
     ||| |||:|||:||| :::::|||| |||
   5 CysGlnSerAlaArgGlyProArgThrProLeuPro........... 17

2247 CTGCAGAGTCCCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTA 2198
     || :::|| |||:|||:
  18 ....HisCysPro...........ProProCysL 24

2197 TA......AGTCCAGCTGAAACCCCTTGGAAGGCCTCCAGTCAGGCAG 2154
     :: :!|||||| ||||||| |||||
  24 euProGlyAlaProAspProGlnThrArgArgLeuProPro....... 37

2153 CCCTAGAGACTGGGAGAGAGAGAGGGACGGCCCCAGCCCCAGCTGTGC 2104
     ||| ::||| ::||| |||||:|:::
  38 ....GlyTrpGlyGlnArgThrAlaProThrInValGlyLe 50

2103 AGCTACGCCACCTCAGCAGCACAGGGTGGCAGCAGAGAGCCACATTACTTT 2054
     ||| |||:::::|
  50 uAlaAspAlaAlaSerProAspGluLeuGlnAspGln............ 62

2053 GGCAGCAACAGAGAACTGGCAGCCAGCCCCATGGGCTAACAG 2004
     |||||| ||| ||||| ||| :::
  63 ....AlaSerGlyAlaArgProGlyGlyGlyAsn 72

2003 GAGCGGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCACCCAATGTGCTG 1954
     ::: ||||||| ::|||||||.::||||||
  73 ArgValGlyAlaGlyArg.......GlyArgProGlyThrProAla.... 85

1953 GAAGTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTTGTCAAATACTAC 1904
     |||||||| |||||||||
  86 ....ProSerArg.......................... 88

1903 CTGTGTAGCAAAGTAAATGGCAACGACCCA.......... 1872
     |!|||||| ||| |||
  89 ....GlnSerArg...ArgThrGlyProAlaGluGlnAlaAspHis 101

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir1:QQBE3		398.50	375.20	5.2e-14	660	BHLF1 protein - human herpesvir
pir2:S21626	-	373.00	345.39	1.1e-12	1453	collagen alpha 1(I) chain prec
pir2:S23809	-	364.00	337.13	3.2e-12	1414	collagen alpha 2(II) chain homo
pir2:S59856	-	358.50	331.71	6.2e-12	1464	collagen alpha 1(III) chain pr
pir1:CGHU7L	-	351.50	325.12	1.4e-11	1466	collagen alpha 1(III) chain pr
pir1:CGHU1B	-	348.00	320.77	2.2e-11	1690	collagen alpha 4(IV) chain pre
pir1:CGBO7S	-	344.50	321.02	3.4e-11	1049	collagen alpha 1(III) chain pr
pir1:CGHU1S	-	343.00	317.14	4.0e-11	1464	collagen alpha 1(I) chain prec
pir1:CGCHLS	-	342.00	318.72	4.6e-11	1042	collagen alpha 1(I) chain - ch
pir2:S18803	-	335.50	308.39	9.8e-11	1843	collagen alpha 1(V) chain - ha
pir2:I48103	-	335.00	309.68	9.9e-11	1549	type VII collagen - Chinese ha
pir2:I50694	-	335.00	313.82	1.0e-10	886	collagen alpha 1(III) chain - c
pir1:CGHU1V	+	333.00	306.06	1.3e-10	1838	collagen alpha 1(IV) chain prec
pir2:JQ2389	+	333.00	315.48	1.4e-10	516	sucrose transport protein - pot
pir2:S23810	-	330.00	304.25	1.9e-10	1603	collagen alpha 1(XVI) chain pr
pir1:CGHU6C	-	328.00	302.93	2.4e-10	1487	collagen alpha 1(III) chain pre
pir2:S28052	+	327.00	309.72	2.9e-10	525	sucrose transport protein - spi
pir2:S43142	+	326.00	308.66	3.0e-10	533	sucrose transport protein - cas
pir1:CGHU2E	-	324.50	299.35	3.7e-10	1546	collagen alpha 2(XI) chain pre
pir1:CGHU2V	-	324.00	299.60	3.7e-10	1496	collagen alpha 2(V) chain prec
pir2:S38196	+	323.50	306.61	4.4e-10	512	sucrose transport protein SUC2
pir2:T00773	+	323.50	306.61	4.4e-10	512	sucrose-proton symporter homolo
pir1:S22917	+	323.00	297.28	4.4e-10	1691	collagen alpha 5(IV) chain pre
pir2:I49607	-	323.00	298.18	4.5e-10	1497	procollagen type V alpha 2 - m
pir2:B41182	-	321.00	296.35	5.7e-10	1487	collagen alpha 1(II) chain pre
pir1:CGBO1S	+	321.00	301.15	5.9e-10	779	collagen alpha 1(I) chain - bov
pir2:S48789	+	319.50	302.93	7.7e-10	507	sucrose transport protein - con
pir2:A41182	-	318.50	294.35	7.7e-10	1419	collagen alpha 1(XI) chain pre
pir1:CGHU1E	-	316.00	290.21	1.0e-09	1806	collagen alpha 1(IV) chain pre
pir2:A54849	-	312.50	283.30	1.6e-09	2944	collagen alpha 1(VII) chain pre
pir2:S16366	+	312.00	287.10	1.6e-09	1763	collagen alpha 2(IV) chain prec
pir1:CGHU2V	-	312.00	287.85	1.7e-09	1496	collagen alpha 2(V) chain prec
pir2:I49607	+	311.50	287.38	1.8e-09	1497	procollagen type V alpha 2 - m
pir2:A45748	-	311.00	290.52	1.9e-09	920	collagen alpha 1(VII) chain - m
pir1:CGRT1S	-	310.00	291.92	2.2e-09	671	collagen alpha 1(I) chain - rat
pir1:A40333	-	309.50	285.52	2.3e-09	1492	collagen alpha 1'(II) chain pr
pir1:CGHU3B	-	309.00	284.22	2.4e-09	1670	collagen alpha 3(IV) chain pre
pir2:B40333	-	308.00	284.14	2.7e-09	1486	collagen alpha 1(II) chain pre
pir1:CGBO7S	+	308.00	286.73	2.8e-09	1049	collagen alpha 1(III) chain pre
pir2:A45407	-	307.00	281.98	3.0e-09	1752	collagen alpha 3(IV) chain - s
pir1:S22917	-	307.00	282.25	3.0e-09	1691	collagen alpha 5(IV) chain pre
pir1:CGHU1D	-	306.50	288.53	3.4e-09	680	collagen alpha 1(X) chain precu

Wed Sep 29 14:26:56 1999

```
                                                     ||||||||
312 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyGly 328

                                                     ||||||||
1008 CAGC........AGTGGGGCGACA.......GCTGGTGGGGCCCA 951
      :::                                           :||||||
328 yAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyP 345

      :::                                CCGACAGCCCTTCTGCTG.....GCTGGTGGGGCCCA 951
983 .........                                   |||
345 roAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluPro 361

950 GCGCTGCCTCCTCAG...............................CCACC 931
     ||||||                :::                       ||||
362 ArgLeuProGlnAspLeuAlaAlaAlaAlaArgCysProAlaGlyProPr 378

930 AGCAGTGTGGCTGCTACGCAGCTGAGGAAGATGAGGGTGAGCAGGCCAAA 881
     |||                         :: |||   :::::::::|||
378 oProThr......................ArgSerGlyAlaAlaAlaGlnA 388

880 GA......GGCACTCCTCCTGGGTGCCCA.......GGTAGGGGGCCAGGG 843
     ||                          ::||||||||
388 rgThrHisArgArgProProGlyCysProArgSerAlaAArgAsnProGly 404

842 CACTGGTGTCCCAGTCAATGGCAGGCAGGAGGTAGCCCAGGCAGCCCCCA 793
     ||||||                    :::
405 ......CysProArgThrTrpArg................              410

792 AGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCCGACAGTGGTCCGG 743
     |||:::|||                      ::|||
411 ......ArgArgSerGlyAlaAlaGlnArgGlyHisProP 421

742 GTCCCGGAAGAGGTCAGAGA...................GCAGGGCCTCCAGTG 708
     ||||||                                 :::|||   |||:||
421 roProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla 437

707 GAGTGAAGCACACCTGCCCACAGAAGTCCAGCAGCCCCACGCCCAGGATG 658
     ||                     ::|||||||||||||      ||||||
438 AlaProGlyAlaProGlyl......ThrProAlaAlaProGlyProGlyl 452

657 AGCAGTGCCAGCTCCA.......GGGGCCTG................GGATCCG 627
     :::||  ::||||    ||||||
452 yGlyAlaAlaValProSerGlyIaThrProHisProGluArgGlySerG 469

626 GGCACAGCAGCCCCTGCTAGCCAGCCCCTTGGGATGAGAAGAGAGGGCTC 577
     :::|||                    ::::::::::::
469 lyProAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGlu 485

576 AGCAGGATGCCCAAGGAC........................          559
     |||:|||:||:||:|||
486 ProArgLeuProGlnAspLeuAlaAlaAlaAlaGlnArgCysProAlaGlyPr 502

558 ..........................AGTGCCCAGATGAAGGGCCGGCGGGGC 531
                              ||||||
502 oProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProP 519

530 CATAGCCTCCACGCGCCAGTGGTCACTGCTGAGCCTGGGACACAG 481
     ||                    :::     ||||||||
519 ro..........................GlyCysProArgSer.AlaArgAs 527

480 ACCAGGCCCAGCACTGACCACCATGGTCATGAACTTCTC 431
     :|:||||                 |||       |||
527 nProGly..........CysProArgThrTrpArgArgSerG 539

430 CTCTA..............CCCCACTTCCAGCAGCAGGTTAGCAGGTTGACC 393
     ||||||:::          ||||||:||  |||
539 lyAlaGlnArgGlyHisProProProGlyAlaGlyGln..........     551

392 TGATGCCTGCGGCCAAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACC 343
     |||||                        :::   |||        ..Pr 561
552 ......ArgProSerGlyProThrGlyGlyArg................    

342 AGCAAGAGCTGGCTTCCGGTGCCCAGCAGGCGGCTCACCCACAGCCT 293
     ||||                 |||:::|||:::||||||
```

```
1871 ........GGCCTGCGGCAGCACCATATAGGCAGTGACA 1840
                  |||   |||:::|||
102 AlaHisSerAsnProThrGlyGlyCysSerAspProGlnArgSerProAr 118

1840 GACTGGCTGAGCTGGACAATGGAGCCCCATAAACAGGATGGGGCCACCTG 1790
      :||:      |||     :::::|||      ..GluG 129
118 gThrArgGlnAlaGlyTyr.............AlaLeuGly 129

1790 GGACGCAGGAAGGCACTATCCGGATGGCGAGGTCCA..........     1752
      ||::||||||      ||| ||||||||
129 lySerAlaGlyLeu.........GlySerArgGlyProArgProHisPro 142

1723 ..........GGCAGATGCCCCGGCCCCGGAACCACCCTG 1723
                   :::|||||       |||
143 AlaPheGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTr 159

1688 ..........CACCCACCACACCACACGTACGGA 1688
               ||||||||||||
159 pArgArgSerGlyAlaGlyGlnArgGlyHisProProPro......      172

1638 GACATCACAGGCAGAGGCCCCGCAGAGCGCGGGTGGAGCTGGGAGCAGGC 1638
      :::|||||||
172 ..........GlyAlaGly 175

1588 CACTGCCTCCAGCACCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTA 1588
      ||||||||||
175 ..........     175

1538 GGGCCTGGCAGGAAGCTCGTCATCAGGCGTGTCTCCACTGCTTAGCACCTCC 1538
      :::||||||
175 ..........     176

1488 AGTGTCCCCTCGGTATTTGGGCAGGAACAACCTGCTTCTCCGGTGTAGGA 1488
      ||| |||  |||:::|||:::
176 lnArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro 192

1438 GGGAGGCCAGTGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCG 1438
      |||  |||::::||      :::||||
193 GlyThrProAlaAlaProGlyProGlyGlyIyAlaAlaValProSer...  208

1392 GTGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGTCGGACAGGC..... 1392
      :::||||||   |||:::   |||:::|||
209 ..........GlyAlaThrProHisProGluArgGlySerGlyProA 221

1356 ..........ATGTGGCACCCGCAGCCCACAGGGAAAGCTCCCACAC 1356
                :::   |||: |||:||
221 laAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArg 237

1306 TGGCCAAATAGACTGCTCGAGTCGCCAAATCGCTGCACCAGCCGGTCCATG 1306
      |||:: ||| |||                ::|||||||||
237 ..........ArgCysProAlaGlyProPr 253

1259 ACCAGAGAGAAGA...CCAGGAGATGGCGCACTGCAGGAACAGCCCCCAG 1259
      ||| |||:||:::      ||||||
253 oProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProG 270

1209 GCTGCCATCCGAACGCCTTCATCATAGTGTCTCCGGGCCTTCGGTGCCCG 1209
      ||||||                     :::|||     ||||||
270 lyCysPro..........ArgSerAlaArgAsnProGlyCysPro 281

1159 GCTCAGCTCTGGGCACGCGCCCTGTACGGCTACAGCCCCAGAAATCCGTG 1159
      ||      |||   |||  ||: |||  |||
282 ...ArgThrArgArgArgSerGlyAlaGlnArgGlyHisProPro.. 296

1109 TAAAACAGCGTGAAGGTCATGAGTCCATCCAGCTGCCAGCTTCAGCCAC 1109
      ||:::||||
296 ..........     296

1059 GAAGAGCCGGCCAGGGTGCGGGGCATGCGGCCAGCCACACCTGGTGCAGCC 1059
      ||||||||||            |||  |||:||
297 ....ProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgPro 311

1009 GGGGAAGCAGGGCGGGCCCCAGGTTCCGGAAAGCCAAGCCAAGCGGCCCGGCCATGGA 1009
```

```
A;Molecule type: DNA
A;Residues: 1-25 <RE2>
A;Cross-references: GB;K01688; NID:g192246; PID:g553881
R;Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741
A;Accession: S39789
A;Molecule type: DNA
A;Residues: 1-80,'E',82-105,'D',107-185;1031-1201,'G',1203-1218,'E',1220-1221,'T',122
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an ind1
A;Reference number: I48300; MUID:94434105
A;Accession: I48300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-80,'E',82-105,'D',107-147 <REF>
A;Cross-references: EMBL:X54876; NID:g50486; PID:g50487
C;Genetics:
A;Gene: COL1A1
A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 10
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple hel
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <VWC>
F;152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

alignment_scores:
      Quality:  375.00              Length:     1037
      Ratio:      0.856             Gaps:         63
Percent Similarity:  42.237    Percent Identity:  27.483

alignment_block:
US-09-030-606-110/rev x S21626

Align seg 1/1  to: to: S21626    from: 1 to: 1453

2594  TGCCTGTGT...........CTCTGTGATGGC........AACAG  2569
         |||                |||||||          |||::
   52  CysIleCysHisAsnGlyThrAlaValCysAspAspValGlnCysAsnGl  68

2568  AAGGACCAACAGGCCACCATCCTGATAAAAGGTAAGAGGGGGTGGATCAG  2519
      ::::    |||::|||||||||||::||    |||||||
   68  uGluLeuAspCysProAsnProGlnArgArg...GluGlyGly.......  81

2518  CAAAAAGACAGTGCTGT.......................  2502
          |||||
   82  ...........CysCysAlaPheCysProGluGluTyrValSerProAsn  94

2501  ...........GGGCTGAGGGGACCTGGTTCTTGTTGTTGCCCCTCAG  2464
                |||:::|:::||||||
   95  SerGluAspValGlyValGluGlyPro....................... 103

2463  GACTCTTCCCCTACAAATAAGTCATATGTTCAAATCCATGGAGGAGTGT  2414
                            |||:|       ||||||||||||||
  104  ...................LysGlyGlyProGlyProG  110

2413  TTCATCCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGT  2364
         ||||||||::: |||::
  110  lnGlyProArgGly.ProValGlyPro....................... 118

2363  TAAGGGGCTTAGAGATGGGAAACCAGGTGACTGAGTTTATTCAGCTCCCA  2314
         |||  |||||||| |||||||   ::||||
  119  .ProGly...ArgAspGlyIleProGlyGlnProGlyLeuProGlyProP  134

2313  AAAACCCT......TCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTA  2270
         |||               |||
  134  foGlyProProGlyProProGlyLeuGlyGlyAsnPheAla  150
```

```
561  oAlaAlaProGly..............AlaProGlyThrProAlaAlaP  573
292  CTGGACCATAGTGGGCCAGGCGGGTAGGGGTAGGGCCGTTCAGGCAC  243
     |||||| :::||| :::||                  ::::||
573  roGlyPro.GlyGlyGlyAlaGlyAlaValProSerGly......AlaTh  586

242  TCCAGAACTGCTTCGTCCGGCTCTGCTCCCAGAAGCTGCGGCCTCCCTC  193
     ||||:::  |||   |||||||||::     |||::|||::|||
586  rProHisProGluArg...GlySerGlyPro.........AlaAspProP  599

192  CTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCCAGGGCGCCCATTCTGCCAG  143
     ||::::::  ::||||||  ::::: |||  |||    ::
599  roAlaAlaAlaArgLeuProGluProGluArg.GlnGluProArgLeuProGl  615

142  CCCTTTGGTGCCGGTCCAGTTTCTCAGCCACCATGCTCAACACCTGCTGCTG  93
     :  ||||||: ||:||  ::|||   :::|::     |||
615  nAspLeuAlaAlaLaGln...ArgCysProAlaGlyProProProThrA  631

 92  TGGGGCACCTCAGTGGGACACGTCTCCATCACTCAGATCCTGGC.......  49
     |||:::    :::||||::: :::||||||
631  rgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysPro  647

 48  CGAGGCGCGGCGGCCTGTCACCCGGA  25
     ||:::|||||  ::||||||
648  ArgSerAlaArg...AsnProGly  654

seq_name: pir2:S21626

seq_documentation_block:
collagen alpha 1(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 20-Mar-1998
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>
A;Cross-references: EMBL:U08020; NID:g470673; PID:g470674
R;Metsaeranta, M.; Toman, D.; de Crombrugghe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A;Reference number: S16176; MUID:91274355
A;Accession: S16374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: EMBL:X57981; NID:g50484; PID:g50485
R;French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403
A;Accession: A23982
A;Molecule type: mRNA
A;Residues: 518-1128 <FRE>
A;Cross-references: EMBL:M14423; NID:g192261; PID:g192262
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A;Reference number: I49559; MUID:91274355
A;Accession: I49559
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 735-1130 <RES>
A;Cross-references: GB:M17491; NID:g192263; PID:g192264
R;Harbers, K.; Kuehn, M.; Dellus, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A;Reference number: I49557; MUID:84170331
A;Accession: I49557
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
406 ..........ProSerGlyPro.GlnGlyProSerGlyProProGlyP 418
                  :::|||           |||||||           |||||||
2269 ACC...........................................CTGAGCCTGGG 2256
            :::|                                      ::|||:::
151 SerGlnMetSerTyrArgGlyTyrAspGluLysSerAlaGlyValSerValPr 167

1434 AGGCGGCTGAAGCTGTCACCAGCGCCACCATGTGGACAGGCATGTGGC 1385
            |||||                   |||   :::        |||:::
2255 TAATCCACCTGCAGAGTCCCCGCATTCCAGTGCATGGAGCCCTTCTGGCC 2206
            :::|||        :::     |||                ::|||
167 oGlyProMetGlyProSer.................GlyProArgGlyL 178

418 roLysGlyAsnSerGlyGluproGly...AlaProGlyAsnLysGlyAsp 433
            |||||           :::           |||
1384 ACCGGCAGCCACAGGGAAAGTCTGCCACACTGGCCAAATAGACTGCTCGAG 1335
            ||||||           |||||::::|||||||
2205 TCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGC 2156
            :::           |||           |||
178 euProGlyProProGlyAlaProGlyPro...GlnGlyPheGln...Gly 192

434 ThrGlyAlaLysGlyGluProGlyAlaThrGly.................. 444
            |||||::                       ||||||
1334 TGCCGAATCGCTGCACCAGCCGGTCCATGCACCAGAGAAGACCAGGGAG 1285
            :::|||   |||   |||                |||
2155 AGCCCTAGAGACTGGGGAGAG........................AGGAG 2130
            ||||||                            ||
193 ProProGlyGluProGlyLyGlyProGlyGlySerGlyProMetGlyProAr 209

445 ...........ValGlnGlyPro..ProGlyProAlaGlyGluGluGlyLyY 457
            :::::                   |||   |||
1284 ATGGCCGCACTGCAGGAACA...........................GCCCCAGGCTGCCCAT 1250
            ||||||                         |||   |||
2129 AGGGACGCCCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCACAGCACAGG 2080
            |||   |||:::                  |||||||
209 gGlyProGlyProPro.................................. 215

457 sArgGlyAlaArgGlyGluProGlyProSerGlyLeuProGlyProProGroG 474
            |||||                         |||   |||
1249 CCGAACGC......................CTTCATCATAGTGTC 1227
            ::|||                         :::|||
2079 GTGGCAGCAGAGAGCCACATTACTTTGGCACAACAGAAACTGGCGGCCA 2030
            |||   |||:::
216 ...................................GlyLysAsnGlyAspAspGlyGlyGlu 223

474 lyGluArgGlyGlyGlyProGlySerArgGlyPheProGlyAlaAspGlyVal 490
            :::|||
1226 TCCGGGCCTCCGTGCCCCGGCTCAGCTCTCGGCACGCCCTGGTACAGCCCC 1177
            :::||||||:::                  ::            |||||||
2029 GCCCGGCAGCCCATGGGGGCTAACAGGAGCCGGGAGCTGGGACCCAGTGA 1980
            |||   ::||||                  |||   ||||||:::|||
224 AlaGlyLysProGlyArgProGlyGluArgGlyLyProProGlyProGlnGl 240

491 AlaGlyProLysGlyProSerGlyGluGlyArgGlyAlaProGlyProAlaGl 507
            :::                         ::            |||||||
1176 TCGCCCACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCA 1127
            ||||||                         |||   |||||||
1979 GGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTTCTACGGTGAGTATTG 1930
            ||||||                         :::
240 yAlaArgGlyLeuProGlyThrAlaGly............LeuProG 252

507 yProLysGlySerPro.................................. 512
            ::|||
1126 GCTCCACAGCTCAGCCACCAAGAGCCGGCGGCAGGGTGCGGGGCATGCGGC 1077
            ||||||
1929 GCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGAC 1880
            :::           |||||||
252 lyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAsp 268

512 ...........................................GlyLysProAlaAsp 512
            ||||||
1076 AGCACAGCTGGTGCAGCCGCGGGGAAGCAGGGGCGCCCAGGT.......... 1038
            :::     |||||||           ||||||||           |||||
1879 CAGACCCAGGCGCTGCGCGGCCAGCAGCACCATATAGGCAGTGA...CAGACTGG 1834
            |||                ::|             |||
269 AlaGlyProAlaGlyProLysGlyGlyProGlySerProGlyGluAsnGl 285

513 GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaAlaLysGlyLe 529
            ||||||||||||           |||||||           |||||
1037 .TCCGGAAAGCCAAGCGGCCCCGGCATGGAAGCAGTCGGGCGCACAAGGA 989
            :::||||:|||:::||||||                |||   |||
1833 CTGAGCTGGACAATGGAGCCCATAAAACAGGGATG........ 1800
            |                         ||||||
285 yAlaProGlyGlnMetGlyPro...ArgGlyLeuProGlyGluArgGlyA 301

529 GlyAlaGlyProGlySerProGlyProAsp......GlyLysThrGlyP 544
            |||           |||   |||           |||:|||||
988 GGGGGCCGACAGCCCTTCTGCTGGCGTCCGTGGGCCCAGCCGCTGCCT..... 942
            |||           |||:::||||||||           |||||||
1799 .....GGGGCCACCTGGACAGCAGGAGGCACTATCCAGGATGGCGAGGT 1755
            ||||||||||||||           ::|           :::||||||
301 rgProGlyProProGlyThrAlaGlyThrAlaAlaArgGly.AsnAspGlyAlaVa 317

544 roProGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProPro 560
            |||||||||           |||||||           |||||||
941 ...........................CCTCAGCCACCAGCAGT 925
            |||           |||   |||           |||||||
1754 CCAGGACAGATGCCCCGGCCCGGACCACCCTGGCCTCGGTGGCTCACCC 1705
            ||||||||:::|||||| ||||| |||   |||           |||
317 lGlyAlaAlaGlyPro.ProGlyProThr...GlyProThrGlyProPro 332

561 GlyAlaArgGlyAlaGlyGlnAlaGlyValMetGlyPheProGlyLysGlyTh 577
            |||||                |||||||
924 GTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCA 875
            ||:::           ||||||           |||||||:::||||||
1704 ACCACCACACGTACGGAGACATCCAGGCAGGAGGCC...CCGCAGCGCC 1658
            ::            |||||||           ||||||:||
333 GlyPheProGlyAlaValGlyAlaLysGlyAlaGluAlaGlyProGlnGlyVal 349

577 rAla...........GlyGluProGlyLysAlaGlyLysAlaGluArgGlyL 589
            :||
874 CTCCT............................................ 870
            :||
1657 GGGTGGAGGT.........GGGAGCAGG......CCACTGCCTCCAG 1626
            :||||:::           |||   :::       |||||:::
349 aArgGlySerGluGlyProGlnGlyValArgGlyGluProGlyProProG 366

589 euProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAla 605
            ||:::                         |||||||
869 ......CCTGGTGCCCCAGGT...AGGGGGCCAGGCACTGGTGTCCCCAG 829
            |||:::||||||                         ||||||:::
1625 CACCCACGTGTCCATTAGGG...AAGGGAGCTCCAGGCTTAGGG...... 1585
            :::||:::|||           |||   |||
366 lyProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAlaAspGlyGln 382

606 GlyAlaGlnAlaProGlyAlaProGlyAlaGlyLysGlyLuArgGlyGl 622
            ||||||:::                         ::GGTAGCCCAG 803
828 TCAATGGCAGGCAGGA................................. 803
            ::|||           |||||||
1584 CCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCCACTCTAGCACTTCCAGT 1535
            |||||| |||           |||||:::||:|||||||
383 ProGlyAlaLys.GlyAlaAsnGlyAlaPro.................. 392

622 uGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaG 639
            ::|||
802 GCAGCCCCCAAGACTGATCATGAAGCACATAGACAGAGTAGGCCTGGCGAC 753
            |||||||:::                         :::||||||
1534 GTCCCCCTCGGTATTTGGGCACGGAAGCACACTGCTTCTCCCGGTGGTAGAGGG 1485
            |||||:::           |||::|||||||:|||||||
393 ......GlyIle..AlaGlyAlaProGlyPheProGlyAlaAlaArgGly 405

639 lyProProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAsp 655
            ||           |||           ::|||||
752 AGTGT...CCGGGTCCC.....GGAAGAGTCAGAGAGAGCAGGGCCTCC 712
            |||           ||||||||:::|||           ||
1484 AGGCCAGTGTGTGTAGGGCAGGATCTGCAGGGGCTGAGAAGGTGAACCCGGTG 1435
```

collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Mar-1999
C;Accession: S23809
R;Exposito, J.Y.; d'Alessio, M.; Solursh, M.; Ramirez, F.
J. Biol. Chem. 267, 15559-15562, 1992
A;Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) c
A;Reference number: S23809; MUID:92348411
A;Accession: S23809
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1414 <EXP>
A;Cross-references: EMBL:M92040; NID:g161435; PID:g161436
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;1207-1414/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

alignment_scores:
         Quality: 364.00   Length: 1224
         Ratio: 0.729   Gaps: 85
Percent Similarity: 40.768   Percent Identity: 26.225

alignment_block:
US-09-030-606-110/rev x S23809

Align seg 1/1 to: S23809 from: 1 to: 1414

3029 GGGAAAGTTGGGGGTAGGGAAAGTTGGGGGTAGGGGAAATTTGGGCAG 2980
     ||| :::||||:: ||| :::||
  43 GlyAspLysGlyGlnLysGlyGluProGlyAspAlaAspPIle...AsnSe 58

2979 TGCCTTCATCAGCCCAGTCCAGAGAGAGTAGAGGG............... 2943
     ||| ::: |||
  58 rAlaAsnPheProProGlyLeuProGlyProValGlyProProGlyProS 75

2942 ...........AGTGGAAGTGGGGGGAACCAGGCT............GGG 2916
                ||||| :||||||
  75 erGlyProSerGlyProAlaGlyAsnAsnGlyProProGlyProAsnGly 91

2915 CCAAGA.........GAAGAGGGTGGTTAGGGAAGCCGTTGAGACC 2878
     |||||| :::||| |||
  92 ProArgGlyAsnProGlyMetAspGlyLeuThrGlyLeuProGlyIlePr 108

2877 TGAAGCCCCACCCTCTACCT.............................. 2858
     ||||||||:::||||| |||
 108 o...GlyProProGly.ProProGlyLysSerGlySerLeuValAlaSer 123

2857 .........TCCTTCAACACCCTAACCTTGGGTAACAGCATTTGGAA 2820
              |||| |||:::   ::
 124 AlaGlnThrSerPheAsnLys.........GlyProSerLeuAlaGl 137

2819 TTATCATTTGGGATGAGTAGAATTTCCAAGGTCCTGGGTTAGGCATTTTG 2770
     ||| :::|||||
 137 yTyr............GlnTyrProGlnAla......GlnAlaAlaG 147

2769 GGGGGCCAGACCCCAGGAGAAGAGAAGAGAATTCTGGCAATGATCAGCCCAATGA 2720
     ||||| ||| |||||
 147 lyThrProGlyProArg................................... 152

2719 CCAGCATCTCAGGGGACCTGATTGTTGGGGATCCCCACCCTACCCAAA 2670
                             |||||||
 153 ....GlyProProGlyProGl 159

2669 TATTAGACACCAACACAGAAAAGCTAGCAATGAATGATTCCTTCTACTTTGT 2620
     : |||::: ||||||
 159 yGlySerArgGlyLysThrGlyLeuThrGlyLysProSerGlyProSer.... 173

2619 TAAATAAAATAAGTAAATATTAAATATTTAAATGCCTGTGTCTCTGTGATGGCAACA 2570
     :   :::   ::::::
 174 ....GlyGluThrGlyProSerGlyAsn 181

2569 GAAGGACCAAACAGGCCACATCCTGATAAAAGTAAGAGGGGGTGGATCA 2520

656 LeuGlyAlaProGlyProSerGlyAlaAlaArgGlyGluArg...GlyPhePr 671
711 AGTGGAGTGAAGCAACCTGGCCACAGAAGTCCAGCAGCCCACGCCCAG 662
671 oGlyGluArgGlyValGlnLnGlyProProGlyProAlaGlyProArgGlyA 688
661 GATGAGCAGTCCA...................GCTCCA 642
688 snAsnGlyAlaProGlyAsnAspGlyAlaAlaLysGlyAspThrGlyAlaPro 704
641 GGGGGCCTGGGATCCGGGCACACAGCAGCCCTGCTAGCCAGCCGCCCTTGGG 592
705 GlyAla......ProGlySerGlnGlyAlaProGlyLeuGlnGlyMetPr 719
591 ATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAAGGG 542
719 oGlyGluArgGlyAlaLaGlyLeuProGly......ProLysGlyAspA 734
541 CCGGGCGGGGCCATAGCCGTCCACGCCCAGTGGCTCACTGAGCCTAGGA 492
734 rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGly 750
491 GCGGGACACAGACCAGGCCCAGCACTGGACCACCAATGCCCAGCACCATGGTC 442
751 AlaAlaArgGlyLeuThrGlyProIle................ 758
441 ATGAACTTCTCCTCTCACCCCACTTCCAGCAGCAGAGGCGGCACATAGGT 392
759 ....GlyProProGlyProAlaGlyIleAlaProGlyGlyAspGlyS 771
391 GATGCCTGCGCCAAACACACCTCCAGGC.................. 363
771 lyGluAlaGlyProSerGlyProProGlyProThrGlyAlaAArgGlyAla 787
362 ........CAAAGGTTAGCAGTTGACCAGCAAGAGAGCTGGGCTTTCCGG 322
788 ProGlyAspArgGlyGluAlaGlyProProGlyProAlaGlyPheAlaGlyl 804
321 T...GCCGCAGCCAGGCGCGGCGTCACCCACCAGCCTCT.............. 291
804 yProProGlyAlaAsAspGlyGlnGlyAlaLysGlyGluProGlyGlyAspT 821
290 ..............GGACCATAGTGGGCGCAGGCGG 270
821 hrGlyValLysGlyAspAlaGlyProAlaGlyPro.AlaGlyProAlaGl 837
269 G...TAGGGCTCAGGGGCGGTTCAGCCACTCAGAACTGCTTCGTCTCG 223
837 yProProGlyProIleGlyAsnValGlyValAlaProGly............P 850
222 GCTCTGCTCCAGAAGCTGCGGCCCTTCCTCCTTGCTGCGCCAACTGCCT 173
850 roLysGlyProArgGlyAlaAlaAlaGlyProPro.GlyAlaThr......... 863
172 AGGAATCAGCCAGGCGCCCATTCTGCCAGCCCATTTGGTGCGCCGGTCCAGC 123
864 .GlyPheProGlyAlaAlaGlyAlaArgValGlyProProGlyProSerGlyA 880
122 TTCTCAGCCACCCATGCTCAACACCTGCTGTCGTGTGGGCACCTCAGT...... 79
880 snAlaGlyProProGlyProValGlyLysGluGluGlyGlyLys 896
78 .......GGGGACAACGTCTCATCACTCAGATCCTGGCCGA........ 46
897 GlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluValGlyPr 913
45 ...........GGCGGCGGCTGTCACCCGGAGCC 22
913 oProGlyProProGlyProAlaGlyGluLuLysGlySerProGlyAla 928

seq_name: pir2:S23809

seq_documentation_block:

```
439 aLeuGlyProAlaGlyProProGlyGlyValGlyGluArgGlyPro.... 454
1807 CAGGGATGGGCCCACCTGGGACAGCAGGA.................. 1779
          ||||||||||||||      :::|||
 455 .....MetGlyProProGlyMetSerGlyAlaProGlyAlaProGlyAla 469

1778 .........AGGCACTATCCAGGATGGCGAGGT............... 1755
          |||   |||||||||||||| ||
 470 LysGlyAspArgGlyLeuProGlyGluArgGlySerAlaGlySerLysGl 486

1754 ....CCAGGCAGATGCCCCGGCCCGGAACCACCCTGGCCTC 1718
            |||||        ||| :::  ||||||
 486 ySerAlaGlyGluSerGlyArgProGlyGluPro.GlyMetProGlyGln 502

1717 GGTGGGCTCACCCACCACCACACGTACGGAGCACATCACAGGCAGAGGCCC 1668
     ||||||||||
 503 ArgGlyLeuThr......................................GlyPr 508

1667 CGCAGAGGCGCGGGTGGAGGTGGGAGCAGCAGCCACTGCCT............. 1630
     |||||||||||||||||||||||| :::
 508 oProGlyLys.GlnGlyArgAspGlyLysProGlyProAlaGlyAlaPro 524

1629 .....CCAGCACCCACGTGTCCATTAGGG.................. AAGGGAGC 1598
          ||:|::||    ::: |||
 525 GlyGluProGlyAsnSerGlyProAlaGlyAlaLaSerGlyGlnArgGlyLe 541

1597 TCCAGGCTTA..................GGGCCT................. 1582
     |||||||||    ||||||
 541 uProGlyLeuValGlyLeuProGlyProGlyLnGlyGlnArgGlyGluArgg 558

1581 .........GGCAGGAAGCTGGTCATCAGGCTGCGTGTCCTCACTGCTAGCACCT 1540
             |||  :::||| ||| :::|||:::|||
 558 lyGluAspGly.GlyGlnGlySerProGlyAlaPro..................... 569

1539 CCAGTGTCCCCTCGTATTTGGGCAGGAACACCTGCTTCTCCCGGTGGTA 1490
            :::|||::|||||||
 570 ......................GlyLeuThrGlyGluProGlyLysA 578

1489 GAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACC 1440
     |||||:::|||      ||||||
 578 rgGlyGluPro............GlyValAlaGly................. 585

1439 CGGTGAGGGGCGGCTGAAGCTGTCACCACGGCACACTGTGGGACAGGCAT 1390
     |||||:::|||  ||| |||:::|||:::|||
 586 ...............ProProGlyProGlnGlySerAlaGlyGl 595

1389 GTGGCACCGGCAG....CCAACAGGGAAAGCTGCCACACTGCCAAATAGAC 1343
          :::||| ||||||||||
 595 uArgGlyAsnGlnGlyProGlnGly............................ 603

1342 TGCTCGAGTGCCGAATCGCTGCACCAGCCGGTCCATGACCAGGAGAGAGA 1293
           :::|||  ||||||||
 604 .....GlnAlaGlySerMetGlyProProGlyPro.............. 613

1292 CCAGGGAGATGGCCACCTGCAGGAACAGCC..........CCAGGCTGCCCATC 1249
     |||||            |||  |||:::|||
 614 ProGlyAlaSerGlyAspAlaGlnGlyAlaGlnGlyAspAsnGlyProProGl 630

1248 CGAACGCCTTCATCATCATATAGTGTCTCCGGGCCGTCCCCGGCTCAGCTCT 1199
     :|||      |||||||||||:::|||
 630 yGlu...............SerGlyProGluGlyPro.............. 637

1198 GGGCACGCCCTGGTACA.........................GCCCCTGCGCCCACGAAAT 1164
     ||||||                |||:::|||
 638 ......ProGlyAlaArgGlyAlaGluArgGlyAlaProGlyGluArgGly 651

1163 CCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCA 1114
     :|||      |||  ::: |||
 652 Pro.....................GlnGlyLeuThrGlyAlaGl 659

1113 GCCACGAAGAGCC...GGCCCAGGGTGCGGGGCATGCGGGCAGCACAGCT.... 1068
     |||||:::  |||:::|||::: |||:::|||
 659 nGlyArgArgGlyIleGlyLyArgAlaGlyAsnTyrGlyAlaThrGlyP 676
```

```
 :::|||||| |||            |||      |||
 182 SerGlyProProGlyProSerGlyLeuProGlyArgProGlySerAspGl 198
2519 GCAAAAGACAGTGCTGT......GGGCTGAGGGGACCTGGTTCCTTGTG 2477
     :::                 |||            |||||||||:::|||
 198 yAspAspGlyThrProGlySerGlnGlnArgGlyLnArgGlyProAlaGlyThrP 215
2476 TGTTGCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAATCC 2427
     |||              ::: |||:::|||:::|||
 215 roGlySerArgGlyThrProGlyMetProGlyAlaProGlyMetLysGly 231
2426 CATGGAGGAGTGTTCATCCTAGAAACTCCCATGCAAGAGCTACATTAAA 2377
     |||   |||      ||| :::|||:::|||:::|||         :::
 232 HisGlnGlyLeu.....ProGlyMetThrGlySerLysGlyGluArgGl 246
2376 CGAAGCTGCAGGTTAAGGGGCTTAGAGATGGAAAACCAGGTGACTGAGTT 2327
     :||:|||      ||||||
 246 yGluGlyGly...GluArgGly...SerAspGlySerProGlyProValGly 261
2326 TATTCAGCTCCCAAAAAACCCTTCCTAGGTGTGTCTCAACTAGGA...... 2282
     |||          :::|||  ||| ||| |||
 262 AlaProGlyProAlaGlyProSerGlyGlnProGlyGluArgGlyArgTh 278
2281 ..................................GGCTAGCTGTTAA 2269
                             |||   ||| |||
 278 rGlyProAlaGlySerGlyGlnGlyAspArgGlyAlaAspGly......AlaT 293
2268 CCCTGAGCCTGGGTAATCCA..............CCTGCAGAGTCCCCG... 2234
     |||||||              ::: |||
 293 hrGlySerGlnGlyProProGlySerThrGlyProAlaGlyAlaProGly 309
2233 ..................CAATTCCAGTGCCATGGAGCCCTTCT...... 2210
                    :::|||||| ||||||
 310 MetProGlyIleSerSerGlyAlaLysGlyAlaAspAlaGlySerProGlyAlaAr 326
2209 .GGCCTCCCTGTATAAGTCCAGACTGAAACCCCCTGGAAGGCCTCCAGT 2161
     ||| |||                   ||||||||
 326 gGlySerPro......................GlyLeuGlnG 333
2160 CAGGCAGCCCTAGAGACTGGGAGAGAGG..........AGAGGG 2126
     ::::::                          |||||||
 333 lyAlaArg..........GlyGluArgGlySerGluGlySerGlnGly 345
2125 ACGCCCCAGCCCCCAGCTGTGCAGCTACGC.............. 2096
     |||||||:::|||
 346 GlnThrGlyProProGlyValProGlyArgAspGlySerAsnGlyAlaLy 362
2095 .............ACCTCAGCAGCACAGGGT...............GGCA 2074
               |||:::|||||||||                    |||
 362 sGlySerAlaGlyProSerGlyAlaGlnGlyThrProGlyPheProGlyA 379
2073 GCAGA...............GAGGCACATTACTTTGGCAGCAAC 2045
     :::|||                     ::::||
 379 laArgGlyProProGlyProAlaGlySerProGlyProAlaGlySerLys 395
2044 AGAAACTGGCGCCAGCCCGCAGCCCATGGGGCTAACAGGAGCGGGGA 1995
     :::                ::: |||:::||||||
 396 GlyAspGlnGlyAsnProGlyGlnProGlyAlaGlnGlyGluSerGlyPr 412
1994 GCTGGGACCC..AGTGAGGCAGGCCCT.......CCACCCCAATGTGCTG 1954
     ||||||||   :::||||||||||      |||||
 412 oLeuGlyProArgGlyGluThrGlyProAlaGlyProPro.......... 425
1953 GAAGTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTTGTCAAATACTAC 1904
     :::||||||||                |||||
 425 .......................................... 425
1903 CTGTGTGTAGCAAAGTAAATGGCGACCAGACCCAGCCTGCGCGCAGACACCA 1854
     ::::::||:::|||||| :::      :::|||
 426 .......GlyAlaGlnGlyGlyGluSerGlyGluArgGlySerArgGlyAl 439
1853 TATAGGCAGTGA....CAGACTGGCTGAGCTGGACAATGGAGCCCCATAAA 1808
     :::|||                  ||| :::||||:::  :::|||
 659 nGlyArgArgGlyIleGlyLyArgAlaGlyAsnTyrGlyAlaThrGlyP 676
```

```
1067 .....GGTGCAGCCGGGGAAGCAGGCGGCCCAGGTTCC..........      1035
         |||  |||||  ||:||| ||||||:::|
 676 roProGlyGlnLysGlyGluMetGlyProProGlyAsnValGlyLeuGln      692

1034 ........GAAAG......CCAAGCGGGCCGGCATGACAGCAGTGG      1000
         |||::|:       ||||||||||||
 693 GlyProProGlyGluLeuGlyProSerGlyPro..........ProGl      705

 999 GGGCAAGGAGAGGGGCCGACAGCCCTTCTGCTGCTCGGTGGGGCCCAG       950
         ||||          ||||||:
 705 yAla......ArgGlyProPro..........GlyProS              713

 949 CGCTGCCTCCTCAGCCACCAGCAGTGTGGCTGCTACCCAGGTGAGGAAGA     900
         ::        |||        |||
 713 erGlySerProGlyProAsp..........GlyPro                 721

 899 TGAGGGTGAGCAGGCCAAAGAGGCACT........                   873
         |||:||||||
 722 AlaGlyAlaGluGlyAlaGlyAspArgGlyProValGlyProMetGlyProSerGl  738

 872 .CCTCCTGGGTGCCCAGGT.AGGGGG.........CCAGGGCACT        839
         ||| || |||||| ||||||
 738 yProSerGlyMetProGlyGluArgGlyAspAlaGlyAspAsnGlyGluProGly..  753

 838 GGGTGTCCAGTCAATGCAGGCCAGGAGGTAGCCCAGGCAGCCCCCAAGAC     789
         ||||||:|||||  :::   |||
 754 ..........ProGlyGlyAlaProGlyGlnArgGlyGlnArgGlyAsp    763

 788 TGATCATGAAGGCCATAGACACAGAGTAGGCCTGGCACAGTGGTCCG.....   744
         ||||||||||:||||
 764 ..........ProGlyAspLeuGlyProGlnGl                    771

 743 .GGTCCCGGAAGAGAGGTCAGAGGAGCAGGCCCTCCAGTGGAGTGAAGCACAC  695
         ::||||||| |||||||||| :::  |||||||
 771 ySerProGlySerProGlyPheAlaGlyLeuProProGlyArgSerGlyAsnP  788

 694 CTGGCCACAGAGTCCAGCAGCCCCACGC.........                 666
         ||||| :::::: ::|||
 788 roGlyProGlnGlyGluLeuGlyProThrGlyAlaArgGlyAlaAlaArgGlyValThrGly  804

 665 ..CAGGATGAGCAGTGCAGCTCCAGGGGCCTCCAGGGATCCGGGCACAGC    619
         |||||        |||||||||
 805 GlyProGly.......ProSerGlyProThrGlyAspProGlyProGlyProGl  818

 618 AGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGAT     569
         ||||||||| |||||||||||||||
 818 nGlyProLeuGlyAlaProGlyGlnGlnGlyLeuGluArgGlyLeuThrGly.   834

 568 GCCCAAGGACAGTGCCCAGATGAAGAGGGCCGGCGGCCCATAGCCGTCCAC    519
         ||||::
 835 ..........ProGln..........                           836

 518 GCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGGACCAGGCCCAGC    469
         |||  |||   |||||||||
 837 ..........GlyGlnGlyGlyGlyProGlyPro..                 844

 468 ACTGGACCAATGCCCAGCAACCATGTCATGAACTTCTCCTCTACCCCCAC     419
         ::: ||||| ::
 845 .IleGlySerLeuGlyLysAlaProGlyAlaGlnGlyProGlyProProThrG  861

 418 TTCCAGCAGAGGCCGCACATAGGTGATGCCTGCGCTGCGCCAAACACACCT    369
         |||:::       |||
 861 lyProSerGlyAsnAlaGly..........SerProGlyGlnPro        872

 368 .........CCAGGCCAAAGGTTAGCAGGTTGACCAGCAAGAGC          334
         |||||||  |||:|||:|:|
 873 GlyAlaArgGlyGlyGlyGln.....SerGlySerProGlyGlnPr        887

 333 TGGGCTTTCCGGTGCCGCAGCA...........GGCGGCT             305
         ||||||||:|       |||||
 887 oGlyLeuAlaGlyArgThrGlyProSerGlyGluArgGlyAspLysGlyGlyA  904


 304 CACCCACCAGCCTCTGGACCATAGTGGG...CCAGGGCGGGTAGGGCTCAGG   258
         ::               |||||||| ||      |||   |||
 904 snAspGlyGlnSerGlyPro.ProGlyProProGlyProAlaGlyProAl    920

 257 GGGCCGTTCAGGCACTCCAGAACTGCTTCGTCTC.....GGCTCTGCTC    214
         ||||:|||||   :::|||  |||   ||||||:::|
 920 aGlyGlnSerGly.......IleLeuGlyLeuAlaGlyLeuGlySerGlyP   934

 213 CAGAAGTCGCGGCCTCCT..........                         194
         |||::||||  |||||||||
 934 rOArgGlyProGlyGlyProAlaGlyLyProProGlyAlaAlaGlySerArg   950

 193 ...CCTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTTCTG   147
         |||:      :::||||::||:|||||||||||
 951 GlyProAlaGlyLysSerGlyAspArgGlySerProGlyAla..........   964

 146 CCAGCCCCTTGGTGCCGTCCAGCTTCTCAGCCATGCCCATGTCAACACCTGCT   97
         :::||||||||||||:::||| |||||
 965 ..........ValGlyProAlaGlyAsnPro......Gly.ProAla      974

  96 GCTGTGGGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTGGC.      49
         :::  ||||| ||:|||||  ::::||  |||
 975 GlyGluAsnGlyMetProGlySerAspGlyAsnAspGlyAlaProGlyPr    991

  48 ......CCAGGGCCGCGGCTGTCACCCGAGCCAGGCCGTGCAGGC          9
         |||:: ::    |||||||||:::||
 991 oGlnGlySerArgGlyGluLysGlyGlyAspThrGlyAlaSerGlyAlaAsnG  1008

   8 TGGTTCCC                                               1
         ||
1008 lySerPro                                             1010
```

seq_name: pir2:S59856

Wed Sep 29 14:26:56 1999

```
alignment_scores:
      Quality:  358.50              Length:     1301
      Ratio:    0.723               Gaps:       80
Percent Similarity: 38.125    Percent Identity: 24.443

alignment_block:
US-09-030-606-110/rev x S59856

Align seg 1/1 to: S59856  from: 1  to: 1464
```

```
3114 GGGGCTGAGATGGACAAAGGCTTG........................... 3091
     ||  :||   ||  |||||
 119 GlyArgAsnGlyAspProGlyLeuProGlyGlnProGlyLeuProGlyPr 135

3090 ....GGAAACCGCACTTTGTGCTTCTGGTCCTGCAGTAGCTCCAAAACAGG 3045
       |||:::|                                :::
 135 oProGlySerProGlyIleCys...........................Glus 144

3044 GTTGTGGAGCTGGTGGGGAAAGTGGGG......................... 3016
      :::  ||||||||   |||:::||||||
 144 erCysProThrGlyGlyGlnAsnTyrSerProGlnPheAspSerTyrAsp 160

3015 ...TAGGGGAAAGTGGGGTAGGGGAAATTTGGCAGTGCCTTCATCA 2969
        :::   ||||||||   |||:::|||
 161 ValLysSerGlyValGlyGlyMetGlyTyrProGlyTyrProGlyProAlaGlyPr 177

2968 GCCCAGTCCTAGAGAGAGTAGAGGGGAGT......................... 2940
     |||||||||
 177 oProGlyProPro.ProGlyProGlySerSerGlyHisProGlySerPro 193

2939 GGAAGTGGGGGAACCAGGCTGGGCCAAGAGACAGAGAGGGGTGGTTAGGGAA 2890
     ||||||  |||  :::  ||| |||  |||  |||
 194 GlySerProGlyTyrGlnGlyProProGlyGlyGluProGlyGlnAlaGly.. 209

2889 GCCGTTGAGACCTGAAGCCCCACCCTCTACCTTCCTTCAACACCCTAACC 2840
     ||:::|||||||   :::||||||:::
 210 .ProAlaGlyProProGlyProProGly........................ 218

2839 TTGGGTAACAGCATTTGGAATTATCATTTGGGATGAGTAGAATTCCAAG 2790
                         ::::||||||||:::::::
 219 .................AlaLeuGlyProAlaGlyProAlaGly 227

2789 GTCCTGGGTTAGGCATTTTGGGGGGGCCCAGACCCCAGGAGAAGAAGATTCT 2740
     ||| |||||  ||:::||||| ||| :::
 228 LysAspGlyGluSerGlyArgProGlyArgProGlyGluArgGlyLeu.. 243

2739 GGGCAATGATCAGCCCAATGACCCAGCTATCTCAGGGACCTGATTGTTGGG 2690
                        ::::||||||||:::
 244 ..........ProGlyProProGlyIleLeuLysGlyProAla......G 254

2689 GATCCCCACCCTACCCAAATATTAGACCACCACAGA........... 2651
     ||  ||| :::||:::||:::::::
 254 lyMet.ProGlyPheProGlyMetLys...GlyHisArgGlyPheAspGl 269

2650 ..................................AAAGCTAGCA 2641
                                      ||:::::|
 269 yArgAsnGlyGluLysGlyGluThrGlyAlaProGlyLeuLysGlyGluA 286

2640 ATGGATTCCCTCTACTTTGTTAAATAAATAAGTAAATATTTAAATGCC 2591
     ||||||:::|||                   |||:::
 286 snGlyLeuProGlyAsp................................ 293

2590 TGTGTCTCTGTGATGGCAACAGAAGGACCAACAGGCCACATCCTGATAAA 2541
     :::  |||||||:::||   |||:::|||
 294 AlaProGlyProMetGlyProArgGlyArgProGlyGlu..........Ar 307

2540 AGGTAAGAGGGGGGTGAT...........CAGCAAAAAGACAGTG 2506
     :||||:||   :::|||||||:::||
 307 gGlyArgProGlyLeuProGlyAlaAlaAlaGlyAlaAlaArgGlyAlaAsnAspGlyA 324

2505 CTGTGGGCTGAGGGGACCTGGTTCTTGTGTTGCCCTCCAGGACTCTTC 2456
     |||
```

```
          ||  |||
 324 laArgGly...................................... 326
2455 CCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGTGTTCATCCT 2406
2406 .............................................. 326
2405 AGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGTTAGGGGGC 2356
 326 .............................................. 326
2355 TTAGAGATGGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCT 2306
     ||||||||::::|||||||                   ||  :::|||
 327 ..SerAspGlyGlnProGly.......................ProProGlyPro 336
2305 ................................TCTCTAGG 2298
                                      :::|||
 337 ProGlyThrAlaGlyPheProGlySerProGlyAlaLysGlyGluValGl 353
2297 TGTGTCTCAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATCCA 2249
     |||                           ||  |||:::|||
 353 yProAlaGlySerProGly...............SerAsnGlySerProG 365
2248 ..................CCTGCAGAGTCCCCGCAATCCAGTGCCATGGAGCCCT 2213
     |||:::  ||   ||:::||||||:::|||
 365 lyGlnArgGlyGluProGlyProGlyHisAlaGlyGlnGlyProGlyPro 381
2212 TCTGGCCTCCTGTATAAGTCCAGACTGAAACCCCCTTGGAGGCCTCCA 2163
                                ::
 382 ProGlyProProGly............................As 387
2162 GTCAGGCAGCCTAGAGACTGGGGAGAG....................... 2135
     :::||||||||||   |||||
 387 nAsnGlySerProGlyGlyLysGlyGluMetGlyProAlaGlyIleProG 404
2134 .........................AGGAGAGGGACGCCCCAGCCC 2114
                             ||   |||:::|||
 404 lyAlaProGlyLeuIleGlyAlaAlaArgGlyProProGlyProAlaGlyThr 420
2113 ........................CCAGCTGTGCAGCTACGCACCTCAGC 2088
                             |||:::  :::
 421 AsnGlyIleProGlyThrArgGlyProSerGlyGlyProGlyLysAsnGl 437
2087 AGCACAGGT.........GGCAGCAGAG 2066
     :::||:::|||     |||::::||
 437 yAlaLysGlyGlyProGlyAlaArgGlyGlyLuArgGlyGlyGluGlySerP 454
2065 ...............CCACATTACTTTGGCAGCAACAGAAACTGGGCGCCAGCCCCGG 2024
                    |||       :::  |||::::||||
 454 roGlyIleProGlyProLysGlyGlyGluAspGlyLys.AspGlySerProGl 470
2023 CAGCCCCATGGGCTAAC.................AGGAGCGGGGAGCTGGGACC 1986
     ::::||||||||     ::||:::  |||
 470 yGluPro..GlyAlaAsnGlyLeuProGlyValAlaAlaGlyAlaArgGlyPr 486
1985 CAGT............................GAGGCAG 1975
     ||||
 486 oSerGlyPheArgGlyProAlaGlyProAsnGlyIleProGlyGluLysG 503
1974 GCCCTCCA.....................CCCCAATGTGCTGGAAGTTTCTACGC 1940
     ||||||||                     |||  ||||||
 503 lyProProGlyGluArgGlyGlyProGlyProAlaGlyPro.....Arg 517
1939 TGAGTATTTGGCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGT 1890
     |||       :::   :::::|||  |||
 518 GlyValAlaGlyValGluProGly.ArgAspGlyThr.Pro.......... 529
1889 AAATGGCGACCAGACCCAGGCCTGCGCAGCACCATATAGGCAGTGACA 1840
     :::||:::||||||      :::||||||
 530 .........GlyGlyProGlyIleArgGlyMetProGlySerProGlyGl 543
1839 GACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATG 1800
     :::||||
```

```
560  yProGlyAsnAspGlyLysProGlyProProGlySerGlnGlyGluSerG
1799 ......GGGCACCTGGACACGACCAGGAAGGCAC..............  1773
560  lyArgProProGlyProSerGlyProArgGlyProArgGlnProGlyVal  576

1772 TATCCAGGATGGCGAGGTCCAGCCAGATGCCCCGGCCCCGGAAC.  1730
577  MetGlyPheProGlyProLysGly.AsnAspGlyAlaProGlyLysAsnG  593

1729 ......CACCCTGGCCTCCGGTGGGCTCACCCACCACACGTAC  1691
593  lyGluArgGlyGlyProGlyGlyProGlyLeuPro.............  604

1690 GGAGACATCACAGGCAGAGGCCCCGCAGAGCGCGGGTGGAGGTGGACA  1641
605  .............GlyProAlaGlyLys.AsnGlyGluThrGlyLyP  615

1640 GGGCACTGCCTCCAGCACCACCCAGTGTCCATTAGGG.............  1606
615  roGlnGlyProProGlyProThrGlyProAlaGlyAspLysGlyAspSer  631

1605 ...............AAGGGAGCTTCCAGGCTTAGGG......  1585
632  GlyProProGlyProGlnGlyLeuGlnGlyIleProGlyThrGlyGlyPr  648

1585 ................................  1585
648  oProGlyGluAsnGlyLysLysProGlyGlyProGlyProLysGlyGluValG  665

1584 ......CCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTG  1549
665  lyAlaProGlyAlaProGly.GlyLysGlyAspSerGlyAlaProGlyGl  681

1548 CTAGCACCTCCAGTGTCCCCTCGGTATT.............TGGGCAG  1514
681  uArgGlyProProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaG  698

1513 GAACACCTGCTTCTCCCCGGTGGTAGAGGG.............AGGCCA  1479
698  lyProProGlyProGlyGluGlyLysGlyProAlaGlyLysGlyProProGlyPro  714

1478 GTGTGTAGGGCAGGATCTGCAGGGCTGAAGGTGAACCCGGTGAGGGCG  1429
715  ProGlyAlaSerGlySerProGlyLeu.GlnGlyMetProGlyGluArgG  731

1428 GCTGAAGCTGTCACCACGGCCACACTGTGGACAGGCATGTGGCACCGGC  1379
731  ly.............................GlyProGly  734

1378 AGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGA  1329
735  SerProGlyProLysGlyGluLysGlyProAlaGlyLysGlyAlaAs  751

1328 ATCGGCTGC.................CCAGCCGGT.........  1311
751  pGlyValProGlyLysAspGlyProArgGlyProAlaGlyProIleGlyP  768

1310 ..CCATGACCAGAGAGAAGAGACCAGGGAGATGGGCCACTGCCAGGAACAGCC  1263
768  roProGlyProAlaGlyGlnProGlyAsp..LysGlyGluGlyGlySer  783

1262 CCAGGCTGC...............CCATCCGAACGCCT  1240
784  ProGlyLeuProGlyIleAlaGlyLyGlyProGlyGluArggl  800

1239 TCATCATAGTGTCTCCGGGCCTCCGTCGCCGGCTCAGCTCTGGGCACCGC  1190
800  yGluHis.........GlyProProGlyProAlaGlyPheProGlyAlaP  814

1189 CTGGTACAGCCCCTCGCCCAGCGAAATCCGTGTAAAACACGCGTGAAGGTCA  1140
814  roGlyGlnAsnGlyGluPro.........................  820
```

```
1139 TGAGTGCCATCCAGCTGCACAGCTGCAGCTCAGCGAAGAGACCCGGCGCAGGGGTG  1090
560  ...............................................  820

1089 CGGGGCATGCGCAGCACAGCTGGTGCAGCCGGGGAAGCAGGGGCGCCCA.  1041
560  .............GGTTCCGGAAAGCCAAGCGGGCCCGGCATGGACAGCAGTGGG  999
821  ..........GGTTCCGGAAAGCCAAGCGGGCCCGGCATGGACAGCAGTGGG  830

1040 .............GGTTCCGGAAAGCCAAGCGGGCCCGGCATGGACAGCAGTGGG  999
830  lyGluLysGlyGluGlyGlyProProGlyProGlyAla.............  841

998  GCGACAAGGAGGGGGCGCGACAGCCCTTCTGCTGCTCGGTGGGGCCCAGC  949
842  ...............GlyProThr..........GlySerSerGlyProAl  850

948  GCTCCTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAGGTGAGGAAGAT  899
850  aGlyProProGlyProGlnGlyVal.....................  858

898  GAGGGTGAGCAGGCCAAAGAGGCACT..........CCTCCTGGGTGCCCAGGT  855
859  ......LysGlyGluArgGlySerProGlyProProGlyThrAlaGly  872

854  AGGGGGCCAGGCACTGGTCTCCCAGTCAATGGCAGCAGGAGGTAGCCCC  805
873  PheProGlyGlyArgGlyLeuProGlyGlyProProGlyAsnAsnGlyAsnPr  889

804  AGGCAGCCCCAAGACTGATCATGAAGGCATAGACACAGATAGGCCTGGCG  755
889  oGlyProProGlyPro...........SerGlyAlaProGlyL  900

754  ACAGTGGTCCG.....................GGTCCCGGAAGAGGT  729
900  ysAspGlyProProGlyProAlaGlyAsnSerProGlyAsnPro  916

728  CAGAGAGCAGGGCCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAGTCC  679
917  GlyIleAlaGlyIleGlyLysProLysGlyAspAlaGlyGlnProGlyLysGlyPr  933

678  AGCAGCCCCACGCCCAGGATGAGCAGTGCCAGCTTCCAGGGGCCTGGGATC  629
933  o...........ProGly..AlaGlnGlyProProGlyIleSerProGlyP  945

628  CG...GGCACAGCCCTGCTAGCCACCCAGCCCCTTGGGATGAGAAAGA  582
945  roLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProPro  961

581  GGCTCAGCAGGATGCCCAAGGACGACAGTGCCCAGATGAAGGGCCGGCGGCGG  532
962  GlyMetProGly..ProArgGlySerProGlyProGlnGlyIleIleLysGl  977

531  CCATAGC.........GTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCGG  488
977  yGluSerGlyLysProGlyAlaSerGlyHis..............AsnG  989

487  GACACAGA.........CCAGGCCCAGCACTGACCAATGCCCAGCCACCATGG  444
989  lyGluArgGlyGlnProGlyPro..........................  996

443  TCATGAACTTCTCCTCACCCCCACTTCCAGCAGCAGAGGCGGCACATAG  394
997  .........GlnGlyLeuProGlyGlnProGlyThrAlaGly.......  1007

393  GTGATGCCTGCGGCCAAACACCTCCAGGCCAAAGTTTAGCAGGTTGAC  344
1008 ...............GluProGlyArgGlyAsnGlySerAspGlyGlnP  1021

343  CAGCAAGAGCTGGCTTTCCGGTGCC.................GCA  315
1021 roGlyArgGlySerProGlyGlyLysGlyAspArgGlyGluAsnGly  1037
```

```
  314 GCAGGCGGCTCACCCACAGCCTTCTGGACCATAGTGGG...CCAGGCGGGT 268
      ::: |||:|||| ||| || ||| || |||  |||||
 1038 SerProGlyAlaProGlyAla.His.ProGlyProProGlyProV 1054

  267 AGGGCTCAGGGGGCCGTTCAGGCACTCCAGAACTGCTTCGTCTCCGGCTCT 218
      ||| ::|||::|||||||       :::   |||
 1054 alGlyProSerGlyLysSerGly.....AspArgGlyGluThrGlyPro 1068

  217 GCT...CCAGAAGCTGCGGCCTCTCCCTTCCCTTGCTGCCGC.......... 182
      |||  |||::::|||  ::::|||     ::|||
 1069 AlaGlyProSerGlyAlaProGlyProAlaGlyAlaArgGlyAlaProGl 1085

  181 ....CAACTGCCTAGGAAATCAGCCAGGC...................... 158
      ||  ||||||||:::
 1085 yProGlnGlyProArgGlyAspLysGlyGluThrGlyGluArgGlySerA 1102

  157 ..........GCCCATTCTCGCCAGCCCCTTTGGTGCCCGTCCAGCTTCT 119
      ::|||    ||| ::: |||||||
 1102 snGlyIleLysGlyHisArgGlyPheProGlyAsnProGlyProProGly 1118

  118 CAGCCCATGTCAACACCTGCTGTCGTGGGGCCACCTCAGT.......... 79
      |||:||||::|||||||||    :::
 1119 .Ser..........ProGlyAlaAlaGlyHisGlnGlyAlaIleGlyLys 1131

   78 ...............GGGGACACGTCTCATCACTCAGATCCT 52
      |||:::::: |||||| |||:: |||
 1131 erProGlyProAlaGlyProArgGlyProValGlyProHisGlyProPro 1147

   51 GGCCGA...GGCGCGGCGCTGTCACCCGGA 25
      ||::: ||||| |||||
 1148 GlyLysAspGlyThrSerGlyHisProGly 1157
```

seq_name: pir1:CGHU7L

seq_documentation_block:
collagen alpha 1(III) chain precursor - human
N;Alternate names: procollagen alpha 1(III) chain
C;Species: Homo sapiens (man)
C;Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 02-Sep-1997
C;Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90
R;Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A;Reference number: S05272
A;Accession: S05272
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1240, 'V',1242-1466 <PRC>
A;Cross-references: EMBL:X14420; NID:g30057; PID:g30058
R;Ala-Kokko. L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A;Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of huma
erences.
A;Reference number: S04642; MUID:89350838
A;Accession: S04642
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1196 <ALA>
A;Cross-references: EMBL:X14420; NID:g30057; PID:g30058
A;Note: the complete sequence is not shown
R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (
A;Reference number: PE0011; MUID:89378752
A;Accession: PE0011
A;Molecule type: DNA
A;Residues: 1-176 <BEN>
A;Cross-references: GB:M26939; NID:g180813; PID:g180814
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
R;Toman, P.D.; Ricca, G.A.; de Crombrugghe, B.
Nucleic Acids Res. 16, 7201, 1988
A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A;Reference number: S01726
A;Accession: S01726
A;Molecule type: mRNA
A;Residues: 1-170 <TOM>
```

```
A;Cross-references: EMBL:X07240; NID:g30060; PID:g30061
A;Note: the authors translated the codon CAG for residue 154 as His
R;Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collag
A;Accession: S04887; MUID:89386015
A;Molecule type: mRNA
A;Residues: 149-163, 'G',164-240, 'D',242-471, 'D',473-487,'L',489,'S',491-613,'Y',615-6
A;Cross-references: EMBL:X15332; NID:g29545; PID:g930045
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleot
R;Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide pept
A;Reference number: A90399; MUID:77134724
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V',169-225,229-232,'P',234-292,'D',294-398 <SEY1>
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-gal
R;Seyer, J.M.
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V',169-225,229-277,'A',279-292,'D',294,'S',296-398 <SEY2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.
Am.J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusu
fspring.
A;Reference number: I51868; MUID:93304430
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 186-194 <MIL>
A;Cross-references: GB:562925; NID:g386425
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the C
A;Reference number: S59511
A;Accession: S59511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID:g1195576; PID:g1195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr
A;Reference number: A90414; MUID:79000343
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675,'N',677-727 <SEY3>
A;Experimental source: liver
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of
A;Reference number: I55349; MUID:91161621
A;Accession: I55349
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:g180815; PID:g180816
R;Seyer, J.M.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from
A;Reference number: A90438; MUID:80199282
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895,'A',897-964 <SEY4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; C
J. Biol. Chem. 265, 17070-17077, 1990
```

A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A;Reference number: A38303; MUID:91009133
A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PID:g180879
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
R;Mankoo, B.S.; Dalgleish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018,'Y',1020-1183,'S',1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:g30053; PID:g30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha (III)-CB9 from ty
A;Reference number: A90446; MUID:81208139
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979,'A',981-984,'PS',987,'QN',990-1096,'P',1098-1152,'AT',1155,'S',1157-
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
A;Reference number: A93551; MUID:85087944
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155,'P',1157-1466 <LOI>
A;Cross-references: EMBL:X01555; EMBL:X01742; NID:g29584; PID:g29584
R;Miskulin, M.; Dalgleish, R.; Kluve-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804
A;Accession: I52393
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: GB:M13146; NID:g180415; PID:g180416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85215505
A;Accession: I79359
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:g180417; PID:g180417
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. F
A;Reference number: A92516; MUID:85157600
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GE
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently G
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine r
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains inte
ation

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein;
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willebrand factor type C repeat homology <VWH>
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;154-167/Region: amino-terminal nonhelical telopeptide
F;168-1196/Region: helical
F;1091-1099/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predi
F;161,1212/Modified site: allysine (Lys) #status predicted
F;263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
F;263/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F;948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted

alignment_scores:
        Quality: 352.50   Length: 1159
        Ratio: 0.770   Gaps: 67
Percent Similarity: 39.517   Percent Identity: 24.935

alignment_block:
US-09-030-606-110/rev x CGHU7L

Align seg 1/1 to: CGHU7L from: 1 to: 1466

3051 AAACAGGGTTGTGGAGCTGGTGGGGAAAGTTGGGGTAGGGGAAAGTTGG 3002
      |||::||| |:|::||||:||||| || ||||| | ::|||
 161 LysSerGlyValAlaValGlyGlyLeuAlaGlyTyrProGlyProAlaGl 177

3001 GGGTAGGGGAAATTTGGGCAGTGCCTTCATCAGCCCAGTCCTAGAGAGA 2952
                           |||::||::||:|||||:|||:||||::|:
 177 yProProGly.........................ProProGlyPro.ProGly 186

2951 GTAGAGGGG..........AGTGAAGTGGGGGAACCAGGCTGGCC 2914
      :::::|| |||||| |||||::: ||
 187 ThrSerGlyHisProGlySerProGlyTyrProGlyGlyTyrProGlyProPr 203

2913 AAGAAGAAGAGGGGTGTTAGGGAAGCCGTTGAGACCTGAAGCCCCACCCT 2864
       | ||| ||||| ||| | ||||||||| :::||||||:
 203 oGlyGluProGlyGlnAlaGly...ProSerGlyProProGlyProProG 219

2863 CTACCTTCCTTCAACACCCTAACCTTGGGTAACAGCATTGGAATTATCA 2814
      :: |||::||||||||||||||
 219 ly.........................GlyIle.ProGlyPheProGlyMetGly 220

2813 TTTGGGATGAGTAGAATTTCCAAGGTCCTGGGTTAGGCATTTGGGGGGC 2764
      :::|| || ::   ::|||:|||||||:::
 221 IleGlyProSerGlyProAlaGlyLysAspGlyLysGlyGluSerGlyArgProGl 237

2763 CAGACCCCAGGAGAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCAGCT 2714
      :|||||| |||::|:||:: |:::   :::
 237 yArgProGlyGluArgGlyLeu.........................ProGlyProProG 249

2713 ATCTCAGGGACCTGATTGTTGGGGATCCCCCACCCTACCCAAAATTAG 2664
      :::::|||||| ||| ||| ....
 249 lyIleLysGlyProAla.....GlyIle.ProGlyPheProGlyMetGly 263

2663 ACACCAACACAGA..........................AAAGCTAGCAATG 2638
        |||||| |||:||::|||
 263 s:.::GlyHisArgGlyPheAspGlyArgArgAsnGlyLysGlyGluThrG 279

2637 GATTCCCTTCACTTTGTTAAATAAATAAGTAAATATTAAAATGCCTGT 2588
      || |||:::   :::||| |||:::
 279 lyAlaProGly.........................LeuLysGlyGluAsnGlyLeu 289

2587 GTCTCTGTGATGGCAACAGAGACCAACAGCCCACATCCTGATAAAGG 2538

```
500 ProGlyGluLysGlyProAlaGlyGluArgGlyAlaProGlyProAlaGl  516
    :::::  ||||||                                     |
1692 ACGGAGACATCACAGGCAGAGGCCCCGCAGAGC......GCGGGTGGAG 1650
516 yProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGlyP  533
    ::|||              ||||::|||
1649 GTGGGAGCAGGCCACTGCCTCCAGCACCCACGTGCCATTAGGAGGAGGA 1600
533 roGlyMetArgGlyMetProGlySerProGlyGlyLySerAspGly  549
    |||  :::|||       |||  ||
1599 GCTCCAGGCTTAGGGCCTGGCGAGGAAGCTGGTCATCAGGCTGTCTCCACT 1550
    ||||||      ||||||:  :::||||||  ||
550 LysProGly...ProProGly.SerGlnGlyGluSerGlyArgPro....  563
1549 GCTAGCACCTCCAGTGTCCCCTCGGTATTTGGCCAGGAACACCTGCTTCT 1500
                                                  GlyProProGlyPro  568
564 ........................................  GlyProProGlyPro  568
1499 CCCGGTGTAGGAGGAGGCCAGTGTGTGTAGGGCAGGATCTGCCAGGGCTGAG 1450
    ||| |||  ||||||:::||                GlyValMetGlyPhe.  580
569 SerGlyProArgGlyGlnPro..........GlyValMetGlyPhe.  580
1449 AAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCACACGGCCACACTGTG 1400
                                                    ;
581 .......................................... ProGlyProLysGlyA 586
                                                  ||| |||
1399 GGACAGGCCATGTGGCACCGGCCAGCCACAGGGAAAGCTGCCACACTGGCCA 1350
    ::  ||||||
586 snAspGly........................  AlaProGlyLys...... 592
1349 AATAGACTGCTCCAGTGCCGAATCGCTGCCACCGGTCCATGACCAGA 1300
    :::  :::  :::::|||::|||:::  |||:::
593 ..........  AsnGlyGluArgGlyGlyProGlyProGlyProGl 605
1299 GAGAGAACCAGGGAGATGCGCACTGCAGA.................. 1269
    :  ||||||:::  ::|||
605 nGlyProProGlyLysAsnGlyGluThrGlyLyProGlnGlyProProGlyP 622
1268 .ACAGCCCAGGCTGCCCATCCGAACGCCTCATCATAGTGTCTCCGGG 1221
    :::  :::::::::|||::|||:::  :::|||
622 roThrGlyProGlyGlyAspLys..............  GlyAspThrGly 633
1220 CCTCGGTGCCCGGCTCAGCTCTGGGCACGCCCTGGTACA......... 1182
    |||    |||  :::  ||||||       ||
634 ProProGlyProGlnGlyLeuGlnGlyGlyThrGlyLyGlyProPr 650
1182 ................................................ 1182
650 oGlyGluasnGlyLysProGlyGluProGlyLyProLysGlyAspAlaGlyA 667
    ||                          |||
1181 ................................  GCCCCTCGCCCACGA 1167
667 laProGlyLeuArgGlyLyLysGlyAspAlaGlyLyAlaProGlyGluArg 683
    |||||                        ||
1166 AATCCGTGTAAAACAGGCTGAAGGTCATGAGTGCCATCCAGCTGCACAGC 1117
    :::|||                    ::  :::  ||
684 GlyPro......................  ProGlyLeuAlaGlyAl 691
1116 TCAGCCACGAAGAGCCGGCCGGCAGGTGCGGGGCATGCGGCCAGCACAGCTG 1067
    ||||:::|||||||                  :::  |||
691 aProGlyLeuArgGlyLyAlaGlyLyProProGluGlyGlyLysG 708
1066 GTGCAGCCGGGGAAGCAGGGGCGCCCAGGTTCC................. 1035
    |||||                  ||
708 lyAlaAlaGlyProProGlyProGlyAlaAlaGlyThrProGlyLeu 724
1035 .................................................. 1035
725 GlnGlyMetProGlyGluArgGlyGlyLeuGlySerProGlyProLysGl 741
    |||||                       :::
1034 .....GGAAAGCCAAGCGGGCCCGGCCATGGACAGCAGTGGGGCGGACA 993
    |||:::|||:::||||||:::|||
741 yAspLysGlyGluProGlyGlyProGlyValAlaAspGlyValProGlyLysA 758
```

```
290 ProGlyGluAsnGlyAlaProMetGlyProArgGlyAlaProGl 306
    :::::  ||||||                      ||
2537 TAAGAGGGGGGTTGGATCAGCAAAAAGACAGCTGCTGTGGGCTGAGGGGACC 2488
    |:::||||||                      ::|||
306 yGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaAargGlyAsn. 322
2487 TGGTTCTTGTGTCGTGCCCCTAGAGACTCTTCCCCTACAAATAAGTCATA 2438
322 .................................................. 322
2437 TGTTCAAATCCCATGGAGGAGTGTTTCATCCTAGAAACTCCCATGCAAGA 2388
322 .................................................. 322
2387 GCTACATTAAAACGAAGCTGCAGTTAAGGGCTTAGA......GATGGGA 2344
                                    |||  ||||||
323 ...........................  AspGlyAlaAargGlySerAspGlyG 331
2343 AACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTG 2294
    ::|||||||                        |||  :::|||
331 lnProGly........................  ProProGlyProPro..GlyThr 340
2293 TCTCAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATCCACCTGC 2244
    :::  :::|||  ||
341 AlaGlyPhePro..............................  GlySerProGlyAl 349
2243 AGAGTCCCCGCCATTCCAGTGCATGGAGCCCT......TCTGGGCTCCCTG 2200
    |||||||             :::  |||  ||||||:::::|||
349 aLysGlyGluValGlyHisAlaGlySerProGlySerAsnGlyAlaProG 366
2199 TATAAGTCCAGACTGAAACCCCCTTGGAGAGGCCTCAGGCAGCCCCT 2150
    |||  ||||||:::|||
366 ly..............................  GlnAargGlyGlyProGlyPro 371
2149 AGAGAACTGGGGAGAG......AGGAGAGGGGACGCCCCAGCCCCCAGCTGT 2106
    ||||:::  :::|||  |||  ||||||:::::
372 GlyProGlnGlyHisAlaGlGlyAlaGlnGlyProProGlyProProGlyIl 388
2105 GCACAGTACGCCACCTCAGCAGCAGCACAGGGTGGCAGCAGAGAGCCACATTACT 2056
    :  :::::::::|||  |||  |||::::::|||
388 e..................  AsnGlySerProGlyGlyGlyLyGlyGlu. 397
2055 TTGGCAGCAACAGAAACTGCGCGCCAGCCCGGCAGCCCCATGGGGCTAAC 2006
    ||||||    |||  :::  :::
398 ..............  MetGlyProAlaGlyIlePProGlyAlaProGly 408
2005 AGGAGCGGGGAGCTGGACCCAGTGAGGCAGGCCCCTCCACCCCAATGTGC 1956
                                  ||||||  |||:::
409 LeuMetGlyAlaArgGlyPro.......  ProGlyProAlaGlyAlaAsnGl 423
1955 TGGAAGTTTCTACGCTG......AGTATTTGGCCCAAGTCGCTCTTGTCAAA 1910
    :::  |||||                  ||:::::|||
423 yAlaProGlyLeuArgGlyValGlyAlaGlyEluProGlyLyLysAsn.... 437
1909 TACTACCTGTGTAGCAAAGTAAATGCGACCAGACCCA.............. 1872
                                ::|||  ||
438 .........  GlyAlaLysGlyGluProGlyProArgGlyGlyEluArg 449
1871 ........GGCCTGCGGCAGCAGACACCATATAGGCAGTGACCAGACTGGCTG 1831
    |||:::::|||               ::|||
450 GlyGluAlaGlyIlleProGlyvalProGlyAlaLysGlyAspGlyLyLy 466
1830 AGCTGGACAATGGAGCCCATAAACAAGGATGGGCCACCTGGGACACCAGCAG 1781
    ||||::  :::  |||  |||:::  |||::||||
466 sAspGlySerProGlyGlyArgProGlyAlaAsnGlyLeuProGlyAlaAlaG 483
1780 GA...AGGCACTATCCAGGATGGCGA.........GGTCCAGGCAGATGC 1743
    ||  |||  ||||||:::::||
483 lyGluArgGlyAlaProGlyPheArgGlyGlyProAlaGlyProAsnGlyIle 499
1742 CCCGGCGCCCCCGGAACCACCCTGGCCCTGCGTGCGGCTCACCCACCACCACGT 1693
    ||||||                         :::  ||
```

```
160 GGGCCCATTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCAG..... 116
    |||    ::: ||||:||||||    |||||||||||:::||
972 GlyProGlnGlyValLysGlyGluSerGlyLysProGlyAlaAsnGlyLe 988

115 ..............................CCCATGCTCAACACCTGCTGCTG 91
                                  ||:|| ::: ||| :||
988 userGlyLeuArgGlyProGlyProGln.GlyLeuProGlyLeuAla 1004

90 GGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTGGC.......... 49
   |||    |||::|| |||::| |||||||    .AGGG
1005 Gly...ThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGl 1020

48 ............................................CGAGGCGCGC 39
                                                |||||
1020 yLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluA 1037

38 GGCTGTCACCCGGAGCC 22
   :: ||||||||||||
1037 snGlySerProGlyAla 1042

seq_name: pir1:CGHU1B

seq_documentation_block:
collagen alpha 4(IV) chain precursor - human
N:Alternate names: procollagen alpha 4(IV) chain
C:Species: Homo sapiens (man)
C:Date: 06-Feb-1995 #sequence_revision 03-Oct-1995 #text_change 26-Feb-1999
C:Accession: A55360; S36854; S28777
R:Leinonen, A.; Mariyama, M.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.
J. Biol. Chem. 269, 26172-26177, 1994
A:Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. C
A:Reference number: A55360; MUID:95014445
A:Accession: A55360
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1690 <LEI>
A:Cross-references: GB:X81053; NID:g574805; PID:g574806
R:Suginoto, M.; Oohashi, T.; Yoshioka, H.; Matsuo, N.; Ninomiya, Y.
FEBS Lett. 330, 122-128, 1993
A:Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen
A:Reference number: S36854; MUID:93374047
A:Accession: S36854
A:Molecule type: DNA; mRNA
A:Residues: 1219-1658,'FE',1661-1690 <SUG>
A:Cross-references: DDBJ:D17391; NID:g440365; PID:g440365; PID:d1004733; PID:g457161
A:Experimental source: whole eye
R:Kamagata, Y.; Mattel, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain
A:Reference number: S28777; MUID:93054733
A:Accession: S28777
A:Molecule type: DNA
A:Residues: 1407-1424,'G',1426-1430,'A',1432-1439,'L',1441-1507 <KAM>
A:Cross-references: GB:L01475; GB:L01476
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w
A:Comment: Prolines and lysines at the third position of the tripeptide repeating uni
    ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A4
A:Cross-references: GDB:132673; OMIM:120131
A:Map position: 2q35-2q37
A:Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands
C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha
    mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeri
    er associations in the interrupted helical domain (with disulfide and desmosine cross
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidne
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydro
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
F:39-61/Domain: amino-terminal nonhelical, NH1 <NH1>
```

```
992 .....AGGAGGGGGGCCGACAGCCCTTCGCTGGCTCGGTGGGGCCCAGCG 948
     ||||||||||
758 spGlyProArgGlyProThr.......................GlyProIle 767

947 CTGCCTCCTCAGCCACCAGCAGTGCTGCTACGCAGGTGAGGAAGATG 898
    ||| :: ||||:||:||:||||      |||::::|||
768 GlyProProGlyProAla........................GlyGlnPr 776

897 AGGGTGAGCAGGCAAAGAGGCACTCCTCCTGGTGCCAGT...AGGG 851
    |||   |||:::  |||::|  |||||     |||||||
776 oGlyAspLysGlyGluGlyGlyAla...ProGlyLeuProGlyIleAlaG 792

850 GGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGAGGTAGCCCAGGC 801
    |||||||||| 
792 lyPro.......................................ArgGlySerProGly 798

800 AGCCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCGACAG 751
    ::: ::: ||| :: |||||||
799 Glu................ArgGlyGluThrGlyProProGlyProAl 809

750 TGGT...CCGGGT..CCCGGAAGAGGTCAGAGAGCAGGGGCTCCAGTGG 707
    :||| ||||||  ||||||||:::|||:||    |||
809 aGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyGlyLysGlyGlyG 826

706 AGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGCCCCAGGATGA 657
    ||    :: ||||||||:||||:::||| ::||:||    ||||||
826 luArgGlyAlaProGlyLysGluLysGlyGlyGlyPro...ProGlyVal 841

656 GCAGTGCCAGTCCAGG.......GCCTGGGATCCGGGCACAGCAGC 616
    ||||||| |||||||        |||||:||:|||||||
842 AlaGlyProProGlyGlySerGlyProAlaGlyProGlyProGlnGlnGl 858

615 CCTGCTAGCCAGCCGCCGCCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCC 566
    ||||||||| || :: |||||||
858 yValLysGly...............GluArgGlySerProGlyGlyP 869

565 CAAGGACAGTGCCCAGATGAAGGGCCGCGGGCCCATAGCGTCCACGCC 516
    ||   |||    |||||||
869 rQ...........................................GlyAla 871

515 AGTGGTCACTGCTCGAGCCTAGGAGGCGGGACACAGACCCAGCCCAGCACT 466
    ||| |||   |||||:||:||||
872 AlaGlyPhe.......ProGlyAlaArgGlyLeuProGlyPro..... 883

465 GGACCAATGCCCAGCACCACCATGGTCATGAACTTCTCCTCTACCCCCACTTC 416
    |||  |||    |||||| |||||||
884 ProGlySerAsnGlyAsnProGlyProGlyProGlyP 895

415 CAGCAGCAGAGGCGGCGACATAGGTGATGCCTGCGCGGCCAAACACACCTCCA 366
    ||
895 roSerGlySer..........................................Pro 899

365 GGCCAAAGGTTAGCAGGTTGACCAGCAAGAGACTGGGCCTTTCCGGTGCCGC 316
    |||||   |||    |||::: ||||| |||| ::||||
900 GlyLys......AspGlyProGlyProAlaGlyAsnThrGlyAlaPr 914

315 AGCAGGCGGCTCACCCACAGCCTCTGGACCATAGTGGG..........G 278
    ||||||| :::|||||||||||
914 oGlySerProGlyValSerGlyPro.LysGlyAspGlyGln 928

277 CCAGGGCCGGTAGGGCTCAGGGGCCCGGTTCAGGGCACTCCAGAACTGCTTCG 228
    |||||| ||||:  |||::: |||||| :::|||
929 ProGlyGluLysGlyGlySerProGlyGlyAlaGlnGlyProPro..... 941

227 TCTCGGCTCTGCTCCA.................................G 211
    ::|||||
942 ......GlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaA 956

210 AAGCTGCGGCCCTCTCCCTTCTGCCGCCCAACTGCCTAGGAATCAGCCA 161
    :::: ||:::||||| ||||||||||||:::|||
956 rgGlyLeuAlaGlyProProGlyMetProGly...ProArgGlySerPro 971
```

```
F;62-1466/Region: interrupted helical
F;94-96/Region: cell attachment (R-G-D) motif
F;145-147/Region: cell attachment (R-G-D) motif
F;189-191/Region: cell attachment (R-G-D) motif
F;310-312/Region: cell attachment (R-G-D) motif
F;724-726/Region: cell attachment (R-G-D) motif
F;785-787/Region: cell attachment (R-G-D) motif
F;989-991/Region: cell attachment (R-G-D) motif
F;1212-1214/Region: cell attachment (R-G-D) motif
F;1467-1690/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1471-1569/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1579-1686/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;47,52,55,57,266,400,460,492,494,668,790,828,1095,1131,1294,1317,1375,1407/Disulfide bo
F;142,669/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1480-1566,1513-1569/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
F;1525-1531,1634-1641/Disulfide bonds: #status predicted
F;1588-1683,1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted
```

```
alignment_scores:
        Quality:    348.50           Length:    1439
        Ratio:      0.681            Gaps:      83
Percent Similarity:  35.580    Percent Identity: 23.628
```

```
alignment_block:
US-09-030-606-110/rev x CGHU1B

Align seg 1/1    to: CGHU1B    from: 1    to: 1690
```

```
2701 CTG...ATTGTGGGGATCCC.........................................  2684
     ||| .:||||||||||||||
 328 LeuGlyLeuValGlyAspProGlyLeuPheGlyLeuIleGlyProLysGl            344

2683 .................................CCACCC..........TACC  2674
                                      ||||||
 344 yAspProGlyAsnArgGlyHisProGlyProProGlyValLeuValThrP  361

2673 CAAATATTAGACACCAACACAGAAAAGCTAGCAATGATTCCTTCTACT  2624
     || :::     :::::  ::::: ||||!|||||::::
 361 roProLeuProLeuLysGlyProProGlyAspProGlyPheProGlyArg  377

2623 ...............................................TTGTTAAA  2616
                                                    ||||||:
 378 TyrGlyGluThrGlyAspValGlyProProGlyProProGlyLeuLeuGl  394

2615 TAAATAAGTAAATATTAAATCCTGTCTCTGTGATGGCAACAGAG  2566
     ::::         |||||:::::::::::::::::
 394 yArgProGlyGlu...........AlaCysAlaGlyMetIleGlyProProG  408

2565 GACCAACAGGCCAC..........ATCCTGATAAAAGTAAGAGGGGGTG  2525
     ||||| |||        ::: |||::: ||:||:
 408 lyProGlnGlyPheProGlyLeuProGlyLeuProGlyAlaGlyIlYle  424

2524 GATCAGCAAAAAGACAGTGCTGTGGCTGAGGGGACCTGGTTCTTGTGTG  2475
 425 ProGlyArgProAspSerAlaProGly...................  433

2474 TTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAATCCCA  2425
 433 .................................................  433

2424 TGGAGGAGTGTTTCATCCTAGAAACTCCATGCAAGAGCTACATTAAACG  2375
 433 .................................................  433

2374 AAGCTGCAGGTTAAGGGCTTAGAGATGGAACCAGGTGACTGAGTTTA  2325
 434 .................................LysProGlyLysProGly  439

2324 TTCAGCTCCCAAAAACCCTTCTCTA...GGTGTGTCTCAACTA...GGAG  2281
     :::|||          |||::::       ||||::: |||||  |||
 440 ...SerProGlyLeuProGlyAlaAlaProGlyLeuGlnGlyLeuProGlyLys  455

2280 GCTAGCTGTTAACCCTGAGCCTGGTAATCCACCTGCAGAGTCCCCGCAT  2231
     ::             ||||||::::: ||||||||||
 455 erSerValIleTyrCysSerValGlyAsnProGlyProGlnGlyIleLys  471

2230 TCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAGACTGAAAC  2181
     ::::::||:        ::: |||::::!||
 472 GlyLysValGlyProProGlyGlyArgGly......ProLysGlyGluLy  486

2180 CCCCTTGGAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGAGAGGA  2131
     |||||||||   ::: ::||||
 486 sGlyAsnGluGlyLeuCysAlaCysGluProGlyPro..........M  499

2130 GAGGGACGCCCAGCCCCCAGCTGCAGCTACCGACCTCAGCAGCACAG  2081
     :::::||    |||||||::: ||| |||:::::|||||||||
 499 etGlyProProGlyProGlyLeuProGlyArg...GlnGlySerLys  514

2080 GGTGGCAGCAGAGCCACCATTACTTTGGCAACAACAGAAACTGGCGGCC  2031
     ||| |||||::||:::     ::: ||||||||
 515 GlyAspLeuGlyLeuProGlyTrpLeuGlyThrLysGlyAspProGlyPr  531

2030 A......GCCCGGCAGCCC.................CATGGGGCTAAC.  2006
     |:|| |||              |||||||||||:::
 531 oProGlyAlaGluGlyProProGlyLeuProGlyLysHisGlyAlaSerG  548

2006 ..................................................  2006
 548 lyProProGlyAsnLysGlyAlaAlaLysGlyAspMetValValSerArgVal  564

2005 ......AGGAGCGGGGAGCTGGGACCCAGTGAGGCAGGCCCCTCCA.....  1967
 548 erTyrGly...........SerProGlyPheProGlyLeuLysGlyGlu  315
```

```
3029 GGGAAAGTTGGGGTAGGGGAAAGTTGGGGTAGGGGAAATTTGGGC..  2982
     |||  ||||||||||||  ||| ||||||
 149 GlyPheProGlyGlyArgGlyAlaLeuGlyProGlyGlyProLeuGlyGlyHi  165

2981 .......................AGTGCCTTCATC..........  2970
                            ||||||||::::
 165 sProGlyGluLysGlyGlyLysGlyAsnSerValPheIleLeuGlyAlaV  182

2969 ...............AGCCCAGTCCTAGAGAGAGTAGAG  2946
                    ::|||  |||     :::
 182 aLysGlyIleGlnGlyAspArgGlyAspProGlyLeuProGlyLeuPro  198

2945 GGGAGTGGAAGTGGGGGAACCAGGCTGGCCAAGA............GA  2908
     ||||| :::::||:::  ||||||||
 199 GlySerTrpProGlyAlaGlyGlyProAlaGlyLeuProThrGlyTyrProGlyGl  215

2907 AGAGGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCCCA........  2868
     |   ||| |||    ||| ||  :||||
 215 uProGlyLeuValGlyLeuProProGlyGlnProGlyArgProGlyLeuLysG  232

2868 ..................................................  2868
 232 lyAsnProGlyValGlyValLysGlyValLysGlyGlnMetGlyAspProGlyGlyGluVal  248

2867 ................CCCTCTACCTTCCTTCAACACCCTAACCT  2839
                     :::::|||:::||||||||
 249 GlyGlnGlyGlySerProGlyProThrLeuLeuValGluProGlyAspPh  265

2838 T.......GGGTAACAGCATTGGAATTATCATTTGGGATGAGTA  2801
     :       ||| ||| :::::::    ||||||| |||
 265 eCysLeuTyrLysGlyLysGlyIleLysGlyIleLeuProGlyMet.Val  281

2800 GAATTTCCA.......................................  2792
     ::|||
 282 GlyLeuProGlyProProGlyArgLysGlyValGlyGluSerGlyIleGlyAlaLy  298

2791 ...AGGTCCTGGGTTAGGCCATTTTGGGGGGCCA.....GACCCCAGGA  2752
     ::::::::  |||||||||  |||           ||||||
 298 sGlyGluLysGlyIleLysProGlyPheProGlyProGlyArgGlyAspProGlyLys  315

2751 GAAGAAGAATTCTGGCAATGATCAGCCAATGACCAGCTATCTCAGGGGAC  2702
     |||        ::::::::::::||!:::
 315 erTyrGly...........SerProGlyPheProGlyLeuLysGlyGlu  327
```

```
846 ySerProGlyAlaProGlyGlyLysGlyGlnProGlyGlyAspValGlyProP 863
                   |||||||
565 LysGlyHisLysGlyGluArgGlyProAsp.....GlyProProGlyPh 579

1468 CAGGATCTGCAGGGCTGAGAAGTGAACCCGGTGAGGGCGGTGAACCTG 1419
      |||||  |||||:::::       |||  |||:::::::::
1966 .CCCAATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAGTCCCTC 1918
      |||  |||||        |||   |||:::::::
863 roGlyProAlaGlyMetLys.............GlyLeuProGlyLeu 874
      |||  |||||        |||:::::::::
579 eProGlyGlnProGlySerHisGlyArg......AspGlyHisAlaGlyL 593

1418 TCACCACGGCCACACTGTGGGACAGGCATGTGCACCGGCCAGCCACAGGG 1369
      |||||||              |||  |||
1917 TTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCCACCA........ 1878
      :::|||||||              |||  |||
863 ProGlyArgProGlyAla.................HisGlyProPro..... 884
      ::|||||
579 ..........GluLysGlyAspProGlyProPro 601

1368 AAAGTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGCAC 1319
      ||||||              |||:: :: ::: ::::|
1877 ..........GACCCAGGCCTGCGCAGCACACCATATAGGCA 1846
      ||||||        ::::
863 .GlyLeuProGlyIleProGlyProPheGlyAspAspGlyLeuProGlyP 901
      ||||||        :::
602 GlyAspHisGluAspAlaThrProGlyGlyLysGlyPheProGlyProLe 618

1318 CAGCCGGTCCA..TGACCAGAGAGAGACCAGGAGATGGGCGCACTGCA 1272
      || ||||||                |||:::|||
1845 GTGACAGACTGCTGAGCTGGACAATGAGCCCATAAACAGGGATGGGC 1796
      ||| ||||||        ::|||  |||:::|||
901 roProGlyProLysGlyProArgGlyLeuProGly.............. 912
      ||| ||||||        :::|||
618 uGlyProProGlyLysAlaGlyProValGlyPro...ProGlyLeuGlyP 634

1271 GGAACAGCCCAGGCTGCCCATCCGAACGCCTTCATCATAGTGTCTCCGG 1222
      ||||||  |||:::::||||||                :::::::||||
1795 CACCTGGGACAGCAGGAAGGCAC.............TATCCA 1767
      ||||||  |||:::||||||
913 ......PheProGlyPheProGlyGluArgGlyLysProGlyAlaGluGl 927
      ||||||  :::|||
634 heProGlyProProGlyGluArgGlyHisProGlyValProGlyHisPro 650

1221 GCCTCGGTGCCCGGTCCAGCTCTGGCACGCCCTGGTACAGCCCCTCGCC 1172
      | ||||||:::: ||||||||
1766 GGATGGCGAGGTCCAGGCAGCAGATGCCCCGGCCCGGAA........... 1731
      ||| ||||||||||
927 y.....CysProGlyAlaLys..GlyGluProGlyGlyLysGlyMetS 941
      ||| |||||:|||
651 GlyValArgGlyProAspGlyLeuLysGlyGlnLysGlyAspThrIleSe 667

1171 CACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGCTGC 1122
                   |
1730 ......................................C 1730
                   |
941 erGlyLeuPro....................................... 944
                   |
667 rCysAsnValThrTyrProGlyTyrArgHisGlyProProGlyPheAspGlyP 684

1121 ACAGCTCACCACGAAGACCCGGCCAGCCAGGGTGCGGGCATGCGGC 1077
      |||          |||:::|  ||||||  |||
1729 CACCCTGGCCTTCGGTG................................... 1714
      |||||  |||:::
945 ......GlyAspArgGlyLeuArgGlyAlaLysGlyAlaIleGlyProPr 959
      |||||  |||:::
684 roProGlyProLysGlyProPheProGlyProGlnGlyAlaProGlyLeuSer 700

1076 ...........................AGCACAGCTGTGCAGCCGGGG 1055
                   |||  |||:::: ||||
1713 ..........GGCTCACCCACCCACCACACGTACGGAGACATC 1682
                   |||  |||:::: ||||
959 oGlyAspGluGlyGluMetAlaIleIleSerGlnLysGlyThrProGlyG 976
                   |||  |||:::: ||||
701 GlySerAspGlyHisLysGlyArgProGlyThr.................. 714

1054 AAGCAGGGGCGCCCCAGGTTCC................GGAAAGCCA 1026
      ||  |||||:::
1681 ACAGGCAGGAGGCCCCGAGAGCGCGGGT................... 1654
      ||||||  |||
976 luProGlyProProGlyAspAspGlyPheProGlyGluArgGlyAspLys 992
      ||||||  |||
715 ....AlaGluIleProGlyProProGlyPheArgGlyMetGlyAspP 730

1025 AGCGGGGCCCGGCATGGACAGCAGTGGGGCGACAAGGAGGGGCCGACAGC 976
      ::: ||||||:::      |||  |||
1653 ........GGAGGT...GGGAGCAGGCCACTG...CCTCCAGCACCC 1621
      |||||  |||:::: ||||||::::|||
993 GlyThrProGlyMetGlnGly............ArgArgGlyGlu...... 1003
      ||||||  |||:::: ||||
730 roGlyPheGlyGlyGluLysGlySerSerProValGlyProProGlyPro 746

975 CCTTCTCGTGGTCGGTGG........................GG 957
      |||:::|||:|||
1620 ACGTGTCCATTAGGG...........AAGGGAGCTCCAGGC........ 1591
      ||||||  |||
1004 ......LeuGlyArgTyrrGlyTyrProProGlyPheHisArgGlyGluProG 1018
      ||||||  |||
747 ProGlySerProGlyValAsnGlyGlnLysGlyIleProGlyAspProAl 763

956 .....................GGCCCAGGCGCTGCCTCCTCAGCCACCAGCAGTG 924
                   ||||||  :::
1590 .......TTAGGG.........................CCTG 1581
                   |||||||
1018 lyGluLysGlyGlnProGlyProProGlyProProGlyProGlyLySer 1034
                   |||||||  :::
763 aPheGlyHisLeuGlyProProGlyLysArgGlyLeuSerGlyValProG 780

923 TGGCTGCTACGCAGG...TGAGGAAGATGAGGGTGAGCAGGCCAAAGAGG 877
      ||  |||              |||  |||||||||
1580 GCAGGAAGCTGGTCATCAGGCTGTCCCTCACTGCTAGCACCTTCCAGTGTCC 1531
      ||  |||        ::: |||
1035 ThrGlyLeuArgGlyPheIleGlyPheProGlyLeuProGlyAspGlnGl 1051
      ||  |||        |||
780 lyIleLys.GlyProArgGlyAspProGlyCysProGlyAlaGluGlyPr 796

876 CACTCCT.........CCTGGTGCCCAGT...................A 854
      ||  ||||||  |||||:
1530 CCTCCGGTATT....................................... 1521
      ||  ||||||
1051 yGluProGlySerProGlyProProGlyPheSerGlyIleAspGlyAlaA 1068
      ||  ||||||
796 oAlaGlyIleProGlyPheLeuGlyLeuLysGlyProLysGlyArgGluG 813

853 GGGGGCCA..................GGGCACTGG............ 837
      |||||||
1521 ...............................................AAGGGAGCTCCAGGC 1521
      |||||||
1068 rgGlyProLysGlyAsnLysGlyAspProAlaSerHisGlyProPro 1084
      |||||||  :::
813 lyHisAlaGlyProGlyIleProGlyGlyLysArgGlyHisSerCysGlu 829

836 ..............................TGTCCCAGTCAATGGCAGGC 817
                   ||||||:::::||||:
1520 ..........................TGGGCAGG 1513
                   |||:::
1085 GlyProLysGlyGluProGlySerProGlyCysProGlyHisPheGlyAl 1101
                   |||  :::
830 ArgGlyAlaProGlyIleProGlyLeuGlyProGlnProGlyTyrProGl 846

816 AGGAGGTAGCCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGA 767
      |:::::||||  |||  |||      |||
1512 AACACCTGCTTCTCCCGGTGTAGAGGGAGGGCCA........GTGTGTAGGG 1469
      |:::::|||||||:::|||  |||||:
1101 aSerGlyGluGlnGlyLeuProGlyIleGlnGlyLeuProArgGlySerProG 1118
```

```
766  GTAGGCCTGGCGACAGTGGT...................CCGGGT... 741
     ::||||||||||                           ||||||
1118 lyArgProGlySerSerGlyProProGlyCysProGlyAsp        1134

740  ........CCCGGA...AGAGGTCAGAGAGACAGGGGCCTCCAGTGGAGTG 703
             ||||||  |||||||  |||||||
1135 HisGlyMetProGlyLeuArgGlyGln............ProGlyGluMe 1147

702  AAGCACACCTGGCCACAGA.........AGTCCAGCAGCCCCA...C 668
     :::   |||||| :::.:          |||||||
1147 tGlyAspProGlyArgGlyLeuGlnGlyAspProGlyIleProGlyP   1164

667  GCCCAGGATGAGCAGTGCCA.............................. 648
     ||||||
1164 roProGlyIleLysGlyProSerGlySerProGlyLeuAsnGlyLeuHis 1180

647  .................................GTTCCAGGGGCCTGGGATCC 628
                                       ||| |||  |||
1181 GlyLeuLysGlyGlnLysLysGlyThrLysGlyAlaSerGlyLeuHisAspVa 1197

627  GGGCACACCAGCCCTGTAGCCACCGCGCCCTTGGGATGAGAAAGAGGCT 578
     ::: ||||||||
1197 lGlyProProGlyProValGlyIleProGlyLeuLysGlyGluArgGlyA 1214

577  CAGCAGGATGCCCAAGGACAGTG.......CCCAGATGAAGGGCCCGGC 537
     :::    ||| |||                  |||||| :::  |||
1214 spProGlySerProGlyIleSerProGlyIleProGlyArgGlyLysLysGlyGly 1230

536  GGGCGGCCATAGGTCCACCGCCAGTGGTCACTGCTGCGGCCTAGGAGCGGGG 487
     |||||||
1231 ProProGlyProSerGlyProSerGlyArgProGlyProAlaGlyAlaTh 1247

486  ACACAGA..............CAGGCCCAGCACTGGACAAT 458
         ||||||||
1247 rGlyArgArgAlaProLysAspIleProAspProGlyProGlyAspPleGlnG 1264

457  GCCCA................... 453
     |||
1264 lyProProGlyProAspGlyProArgGlyAlaProGlyProProGlyLeu 1280

452  ....................GCACCATGGTCATGAACTTCTCC 430
                         ||  :::  ||
1281 ProGlySerValAspLeuLeuArgGlyLeuArgGlyGluProGlyAspCysGlyLeuPr 1297

429  TCTACCCCACTTCCAGCAGCAGAGAGGCGCACATAGGTAGGCAGTGATGCCTGCCGC 380
     |  ||||||| ||  |||||
1297 oGlyProProGlyProProGly..........................P 1305

379  CAAACACAACCTCCAGGCCCAAAGG.................TTAGCA 351
     ||   ||||||||  ::
1305 roProGlyProProGlyTyrLysGlyLeuProGlyProProGlyCysAspGlyProArgGlyLysAsp 1321

350  GGTTGACCAGCAAGAGCTGGGCTTTCCGGTGCCGCAG...CAGGCGGCTC 304
     |||  :||||||||  |||||
1322 Gly.GlnLysGlyProMetGlyProPheProGlyLysGlyProHisGlyP 1338

303  ACCCAGCCTCTGGACCATAGTGGG...CCAGGCGGGTAGGGCTCAGGG 257
     :::   ||||||| |||||||
1338 heProGlyLeuProGlyLeuLysGlyGlyLeuProGlyProProGlyArgGlyLys 1354

256  GGGCGTTCAGGCCACTCCAGAAACTGCTTCCTCTGCTCTCCAGAGC 207
     |||   ::|::::::
1355 GlyProThrGlyLeuPro.................GlyProArgGl 1364

206  TGCGGCCCTCTCCCTTGC..........TGCCGCCAACTGCCTAGGA 169
     |||   ::||||||||     ::::|||::: |||
1364 yGluProGlyProProAlaAspValAspAspCysProArgProArgIlePro.Gly 1380

168  ATCAGCCAGGCGCCCATTTCTGCCACGCCCTTTGGTGCCCGGTCCAGCTTCT 119
     :::   ||||||||    ::||
1381 LeuProGlyAlaProGlyMetArgGlyPro................... 1390
```

```
118  CAGCCCATGCTCAACACCTGCTGCTGTGGGCACCTCAGTGGGGACACGT 69
     ::|||:::|||:||  ::||  |||  |||||||
1391 ........GluGlyAlaMetGly...LeuProGly...... 1398

68   CTCATCACTCAGATCTCCGCCGAGGCGCCGGCGTCACCCGAGCCAGC 19
     ::::: ||||||  :||::||::||||   :::
1399 ..MetArgGlyProProGlyCysLysGlyGluProGlyLeuAsp 1414

18   GGCTGCAGG 10
     :::  |||
1415 GlyArgArg 1417
```

seq_name: pirl:CGBO7S

```
F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
F:1-14/Region: amino-terminal nonhelical telopeptide
F:15-1040/Region: helical
F:587-589/Region: cell attachment (R-G-D) motif
F:752-754/Region: cell attachment (R-G-D) motif
F:875-877/Region: cell attachment (R-G-D) motif
F:878-880/Region: cell attachment (R-G-D) motif
F:935-937/Region: cell attachment (R-G-D) motif
F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide
F:95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
F:107,950/Modified site: allysine (Lys) #status predicted
F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental
F:1040,1041/Disulfide bonds: interchain #status predicted

alignment_scores:
     Quality:   348.00            Length:   1139
     Ratio:      0.740            Gaps:       73
Percent Similarity:  41.264     Percent Identity:  26.163

alignment_block:
US-09-030-606-110/rev x CGBO7S

Align seg 1/1 to: CGBO7S from: 1 to: 1049
```

```
3051 AAACAGGGTTGTGGAGCTGGTGGGGAAAGTTGGGGTAGGGAAAGTTGG 3002
     |||:|||  ::||||||||  :::   ||  ::||||
   8 LysSerGlyValAlaGlyGlyGlyIleAlaGlyTyrProGlyProAlaGl 24

3001 GGGTAGGGGAAATTTTGGGCAGTGCCTTCATCAGCCCAGTCCTAGAGAGA 2952
     |||:||||            |||:::|||
  24 yProProGly...............ProProGlyPro.ProGly 33

2951 GTAGAGGGG...........AGTGGAAGTGGGGGGAACCAGGCTGGGCC 2914
     ::::||||  |||:: |||   ||||  ||
  34 ThrSerGlyHisProGlyAlaProGlyAlaProGlyThrGlnGlyProPr 50

2913 AAGAGAAGAGGGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCCCACCCT 2864
     |   ||  |||  ||| |||   |||  ::::|||||:
  50 oGlyGluProGlyGlnAlaGly...ProAlaGlyProProGlyProProG 66

2863 CTACCTTCCTTCAACACCCTAACCTTGGGTAACAGCATTGGAATTATCA 2814
     ||| |||||||:||                       :::
  66 ly.....................................Ala 67

2813 TTTGGGATGAGT...................AGAATTTCCAAGGTCCT 2785
     ::|||| |||                     ||| |||
  68 IleGlyProSerGlyLysAspGlyGluSerGlyArgProGly.ArgProG 84

2784 GGGTTAGGCATTTGGGGGGCCAGACCCCAGGAGAAGAGATTCTGCAA 2735
     |||||| ||| ||||||
  84 lyProArgGlyPheProGlyPro................. 91

2734 TGATCAGCCCAATGACCAGCTATCTCAGGGACCTGATTGTTGGGGATCC 2685
     |||::::::::|||                     :::
  92 ............ProGlyMetLysGlyPro.......AlaGlyMe 100

2684 CCCACCCACCCAAATATTAGACACCAACACAGA................ 2651
     ||| :::::||||:      ||||||
 100 tProGlyPheProGlyMetLys...GlyHisArgGlyPheAspGlyArga 116

2650 ........AAAGCTAGCAATGAATGCCTTCCTTCACTTTGTTAAATAAATAA 2609
     ||||::::|| |||  |||:::
 116 snGlyLysGlyGluProGlyAlaProGly.................... 126

2608 GTTAAATATTTAAATGCCTGTGTCTCTGTGATGGCAACAGAAGGACCAAC 2559
     ::::|||     |||:: :::|| |||||   |||||
 127 LeuLysGlyGluAsnGlyLysValProGlyGluAspGlyLysAlaProGlyPheMe 143

2558 AGGCCACATCCTGATAAAAGGTAAGAGGGGGTGGATCAGCAAAAAGACA 2509
     |||  |||  :::|||
 143 tGlyProArgGlyAlaProGlyGluArgGlyAlaProGlyLeuProGlyA 160
```

```
2508 GTGCTGTGGCGGTGAGGGACCTGGTTGTTCTTGTGTTGTTGCCCTCAGGACTC 2459
     ::|||  |||  :  |||:|||
 160 laAlaGlyAlaAlaGlyAsn................................ 166

2458 TTCCCTACAAATAAGTCATATGTTCAAATCCATGGAGGAGTGTTTCAT 2409
     ||||||||:||                        ||||||||||:
 166 ................................................... 166

2408 CCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAGG 2359
     |||||| |||||  ::::||: :::::|||||          ||
 167 ...........................................AspG 168

2358 GGCTTAGA.......GATGGGAAACCAGGTGACTGAGTTATTCAGCTCCC 2315
     || |||  |||||:::||||||                   ||||||||:
 168 lyAlaArgGlySerAspGlyGlnProGly.................Pro 178

2314 AAAAACCCTTCTCTAGGTGTGTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     :::|||  |||:::   ||:::||
 179 ProGlyProPro...GlyThrAlaGlyPhePro.............. 188

2264 GAGCCTGGGTAATCCACCTGCAGAGTCCCGATTCCAGTGCATGATGGAGCC 2215
     |||::||||   |:|||  ::|  |||||||| ||||  ||||
 189 ......GlySerProGlyAlaLysGlyGluValGlyLysGlySerP 203

2214 CTTCTGGCCTCCCTGTATAAGTCAGACTGAAACCCCTTGGAAGGCCTC 2165
     |||:||
 203 roGly........................................ 204

2164 CAGTCAGGCAGCCCTAGAGACTGGGGAGAG.................AGGAG 2130
     :::||||||:|||  :::   ||||||
 205 SerSerGlyAlaProGlyGlnArgGlyGluProGlyInGlyHisAl 221

2129 AGGGACCGCCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGG 2080
     |||::|||  ||||||                ::: |
 221 aGlyAlaProGlyProPro...............GlyProProG 231

2079 GTGGCAGCAGAGACCCACATTACTTTGGCAGCAACAGAAACTGGCGGCCA 2030
     |||::|||   |||   |||   ::::|||||
 231 lySerAspGlySerPro.......GlyGlyLysGlyGluMetGlyPro 244

2029 GCCCGGCAGCCCATGGGGCTAACAGGAGCGGGGAGCTGGGACCAGTGA 1980
     ||| ::: :::   |||||| |||||
 245 AlaGlyIleProGlyAlaProGlyLeuIleGlyAlaArgGlyPro...... 259

1979 GGCAGGGCCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGGTATTTG 1930
     |||||||                  :::  |
 260 .ProGlyProProGlyThrAsnGlyValProGlyGlnArgGlyAlaAlaG 276

1929 GCCAAGTCGCTTGTCAAATACTACCTGTGTAGCAAATAAATGGCCAC 1880
     :::     ||||||||  ||::::|||||||
 276 lyGluProGlyLys...............AsnGlyAlaLysGlyAsp 286

1879 CAGACCCAGGCCTGCGGCAGACAACCATATAGGCAGTGACGACTGGCTGA 1830
     |||   |||  :::||||
 287 ProGlyProArgGlyGluArgGlyGluAlaGlySerProGlyIleAlaGl 303

1829 GCTGGACAATGAGCCCCATAAACAGGGATG.................. 1800
     :::::|||
 303 yProLysGlyGluAspGlyLysAspGlySerProGlyGluProGlyAlaA 320

1799 ..GGGCCACCTGGGACAGCAGGA...AGGCACTATCCAGGATGGCGAGGT 1755
     ||:||  :::||||||||:: |||||||||:||||:
 320 snGlyLeuProGlyAlaAlaGlyGlyValProGlyPheArgGly 336

1754 CCAGGCAGATGC...........CCCGGGCCGGAA...CCACCCTGGCCCTCG 1717
     :::|||  |||                ::: ||||
 337 ProAlaGlyAlaAsnGlyLeuProGlyGluLysGlyProProGlyAspAr 353

1716 GTGGGCTCACCCACCACCACACGTACGGACACATCACAGGCAGGAGGCCCC 1667
     ||| ::: ||||||  |:::  |||||||| |||
 353 gGlyGlyProAlaGlyProArg.GlyValAlaGly.GluProGl 369

1666 GCAGAGCGCG.......GGTGGAGGTGGGAGCAGCCCACTGCCCTCCAGCAC 1623
```

```
618 lyGlnProGlyAspLysGlyGluSerGlyAla...ProGlyValProGly 633
    ::::::::      |||||||    |||     |||   ::::|
369 yArgAsnGlyLeuProGlyGlyProGlyLeuArgGlyIleProGlySerP 386

854 ..AGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGAGGTAG 808
    |||||||                                      ||||||
1622 CCACGTGTCCATTAGGGAAGGGAGCTCCAGCTTAGGGCCTGGCAGGAAG 1573

634 IleAlaGlyPro..........................ArgGlyGl 640
    |||||||                                ||||||
386 roGlyLeuGlyProGlySerAsnGlyLysProGly...ProProGly.SerGl 401

807 CCCAGGCAGCCCCCAAGACTGATCATGAGGCATAGACAGAGTAGGCCTG 758
    :|||||||:|||              |||    ::|||||:::    |||
1572 CTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCAGTGTCCCTCGGTA 1523

640 yProGlyGlu...........ArgGlyGluGlnGlyLysProProG 651
    :||||||:|||         |||
407 nGlyGluThrGlyArgPro.......................... 407

757 GCGACAGTGGT...CCGGGT...CCCGGAAGAGGTCAGAGAGCAGGCCT 714
    || ::|||||| ||||||| ||||||   :::  |||
1522 TTTGGGCAGGAACACCTGCTTCCCCGGTGGTAGAGGGAGGCCAGTGTGT 1473

651 lyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyAla 667
    ||| :::||||||||||  ||||| ::|||
408 ........GlyProProGlySerProGlyIleProArgGlyGlnProGlyVal 421

713 CCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGCCC 664
    ||| :::||||||||:::::::::|    :::::|||
1472 .......................................AGGGCAGG 1465

668 LysGlyGluArgGlyAlaProGlyGluLysGlyGluGlyProGly...Pr 683
    ||| :::||||||||:||:::  :::::|||
1472 ............................................ 

663 AGGATGAGCAGTGCCAGCTCCAGGG.......GCCTGGGATCCGGGCA 623
    ||:|||    ||| |||   ::       ||||||||||||||
422 MetGlyPheProGlyProLysGlyAsnAspGlyAlaProGlyLysAsnGl 438

683 oGlyAlaAlaGlyProAlaGlyGlySerGlyProAlaGlyProGlyP 700
    ||: :: :::|||||||||||:||||||||||||:|
1464 ATCTGCAGGGCTGAGAAGTGAACCCGTGAGGCGGCTGAGCTGTCAC 1415

622 CAGCAGCCCTGCTAGCCAGCCGGCCCCTTGGGATGAGAAAGAGGCTCAGCA 573
    |||||||||||
438 yGlu.ArgGlyGlyProGlyProGlyProGlyProAlaGlyLys 454

700 roGlnGlyValLysGly......................GluArgGlySerPro 710
    ::: :::
1414 CACGGCCACACT.....GTGGGACAGGCATGTGGCACCGGCAGCCACAG 1371

572 GGATGCCCAAGGACAGTGCCCAGATGAAGGGCCGGC......GGCGGCCA 529
    :::|||||||                           ||||||
455 AsnGlyGluThrGlyProGlnGlyProProGlyProThrGlyProSerGl 471

711 GlyGlyProGlyAla.................AlaGlyPheProGlyGlyAr 722
    ||| ::: ||||||||||:::|||
1370 GGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTGC 1321

528 TAGCGTCCACGCCAGTGTCACTGGCTGGACGGGACACAGAC 479
    ||||||||||||||||||||
471 yAspLysGlyAspThrGlyProProGlyProGlyLeuGlnGlyLeu. 487

722 gGly........................................ProP 725
    :::
1320 ACCA......GCCGGTCCATGACCCAGAGAGA........AGACCAG 1289

725 roGlyPro..........................ProGlySerAsnGlyAsnPro 734
    |||||||                           ||| |||
488 .ProGlyThrSerGlyProProGlyGluAsnGlyLysProGlyGluProG 504

428 CTACCCCCA.............CTTCCAGCAGCAGAGGGCGGCACATAGGTG 391
    ||||||                |||
735 GlyProProGlySerSerGlyAlaProGlyLysLysAspGly.......... 747

478 CAGGCCCAGCACTGGACCAATGCCCAGCACCATGGTCATGAACTTCTCCT 429
    ||||||||                    :::||||
1288 GGAGATGGCCCACTGCAGGAACAGCCCAGGCTGCCCATCCGAACGCCTT 1239

390 ATGCCTGCGGCCAAACACACCTCCAGGCCAAAGGTTAGCCAGGTTGACCAG 341
    ||| |||||||||:::|||   ||  :::
1238 CATCATAGTGTCTCCGGGGCTCGGTGCCCGGCTCAGCTCTGGGCACGCCC 1189

748 ..........ProProGlyProGlySerAsn......GlyAlaProG 759
    ||| |||||||
1188 TGGTACAGCC.....CCTCGCCCACGAAATCCGTGTAAAACAGCGTGAA 1145

340 CAAGAGCTGGCTTCCGGTGCCCGCAGCAGGC................ 309
    :: |||||||
530 oGlyAlaGlyLyMetProGlyGluArgGlyGlyProGlyGlyProGlyP 540

759 lySerProGlyIleSerGlyGlyAspSerGlyProProGlyProGlyGlu 775
    ||||| :::
1144 GGTCATGAGTGCCATCCAGCTGCACAGCTCAGCCACGAAGAGCCGGCGCA 1095

308 ..GGCTACCCCCAGCCTCTGGACCACCATAGTGGG...CCAGGCGGGTAGG 265
    ::::||||||      :::::||||||
541 ...............GlyAlaGlyProProGlyProGlyGlu 549

776 ArgGlyAlaProGlyProGlnGlyPro.ProGlyAlaProGlyProLeuG 792
    ||:::|||:::||||||||  ::||:||||
1094 GGGTGCGGGCATGCGGC..........AGCACAGCTGGTGCAGC 1060

264 GCTCAGGGGCCGGTCAGCACTCCAGACTGCTGCTTCGTCTCGGCTCTGCT 215
    || ::::||        :::||||||  |||
1059 CGGGGAAGCAGGGCGCCCAGGTTCC...GGAAAGCCAAGCGGGCCCGGC. 1014

792 lyIleAlaGlyLeuThrGlyAlaAlaGlyThrAlaGlyThrAlaGlyGlu 808
    ||| |||||||:::|||||||||||||||||
1013 ..................ATGGACAGCAGTGGGGGCGGACA........ 993

214 CCAGAAGCTCGGGCTCTCCT.................... 194
    ||| ||| ::::|||||
809 ProGlyAlaArgGlySerProGlyIleLeuLysGlyGluAsnGl 825

193 ..........CCTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCC 154
    ::::|||         ||||||
992 roLysGlyAspLysGlyGluProGlySerSerGlyValAspGlyAlaPro 599

825 yLysProGlyProSerGlyGlnAsnGlyGluArgGlyProGly.... 840
    |||||||                              |||||||
954 .........AGGAGGGGGCCGACAGCCCCTTCTGCTGGCTCGGTGGGGG 955

153 ATTTCTGCCAGCCCTTTGGT...............GCCGGTCCAGCTTCTCAG 116
    ||| :::|||||         |||:::|||||
600 GlyLysGlyAlaAlaGlyLysProProGlySerAlaGlyThrPr 566

841 ..........ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlu 853
    ||| ::::|||
1059 .......... 

115 CCCATGCTC......AACACCTGCTGCTGTGGGGCACCTCAGTGGGGACA 72
    ||| ::: |||                             |||
954 CCCCAGGCCTGCTCCTCACCACCAGCTGGCTGCTACGCAGGTGAG 905

854 ProGlyArgAspGlyAsn.ProGlySerAspGlyLeuProGlyArgAspG 870
    ||:::: ||| ||:::: |||||||
904 GAAGATGAGGTGAGCAGCCAAAGAGGCACTCCTCCTGGTGCCCAGGT 855
```

```
      71 CGTCTCATCACTCAGATCCTGCCGAGGCGCCGGCGGCGCTGTCACCGGAGCC  22
         :::         ||||||   :::  ||||||||||||   |||||||||||
     870 lyAlaProGlyAlaLysGlyAspArgGlyGluAsnGlyGlySerProGlyAla  886
```

seq_name: pir1:CGHU1S

seq_documentation_block:
collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 26-Feb-1999
C:Accession: I60114; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11
5269; A29439, I53466; A02852; I37247
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five e
A:Reference number: I60114; MUID:88329734
A:Accession: I60114
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-369,'L',371-589 <DAL>
A:Cross-references: GB:M20789; NID:g179593; PID:g179594
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock
Biochem. J. 253, 919-922, 1988
A:Molecule type: mRNA
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human
A:Residues: 1-472 <TRO>
A:Cross-references: EMBL:X07884; NID:g30015; PID:g30016; GB:M36546; NID:g190227; PID:g55
A:Note: submitted to the EMBL/GenBank/DDBJ databases by Prockop, D.J., 13-JUN-1988
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;
Nature 310, 337-340, 1984
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58,'Q',60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:g35657; PID:g35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Ploy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 262, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh
A:Reference number: A39943; MUID:88097389
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:g180876; PID:g553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A:Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A:Reference number: I55237; MUID:85130970
A:Status: translation not shown; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:g180387; PID:g180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control o
A:Reference number: A39943; MUID:88097389
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <CH>
A:Cross-references: GB:M10627; NID:g180383; PID:g553226
R:Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A:Title: In vivo and in vitro noncovalent association of excised alpha(I) amino-termina
rome, type VII.
A:Reference number: A35233; MUID:90202908
A:Accession: A35233
A:Molecule type: protein

A:Residues: 33-52 <WIR>
A:Note: this propetide fragment remained non-covalently bound to a defective, unclea
R:Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.G.; Cole, W.G.; Chan, D.; Bateman, J.
EMBO J. 8, 1705-1710, 1989
A:Title: A base substitution in the exon of a collagen gene causes alternative splici
A:Reference number: S09400; MUID:89356643
A:Accession: S09400
A:Molecule type: mRNA
A:Residues: 156-183 <WEI>
R:Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alp
A:Reference number: A90567; MUID:71038625
A:Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A:Accession: B90567
A:Molecule type: protein
A:Residues: 162-198,'Z',200-201,'Z',203-206,'Z',208-209,'Z',211-228,'B',230,'BB',233,
A:Experimental source: skin
A:Note: evidence for 170-allysine
R:Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller
Eur. J. Biochem. 192, 153-159, 1990
A:Title: A critical crosslink region in human-bone-derived collagen type I. Specific
A:Reference number: S11372; MUID:90382436
A:Accession: S11372
A:Molecule type: protein
A:Residues: 175-187;274-287,'P',289 <BAE>
A:Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
R:Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonza
J. Biol. Chem. 266, 21827-21832, 1991
A:Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen ch
cooperative melting of intact type I collagen.
A:Reference number: I55342; MUID:92042092
A:Accession: I55342
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 258-268:1347-1357 <DEA>
A:Cross-references: GB:S67495; NID:g239007; PID:g239008
A:Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg rep
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagen
A:Reference number: A92069; MUID:71001508
A:Accession: A92069
A:Molecule type: protein
A:Residues: 263-268 <MOR>
A:Experimental source: skin
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R:Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A:Title: Segmental amplification of the entire helical and telopeptide regions of the
A:Reference number: S15989; MUID:90326017
A:Accession: S15989
A:Molecule type: mRNA
A:Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>
R:Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries,
Connect. Tissue Res. 29, 1-11, 1993
A:Reference number: I52905; MUID:93339042
A:Accession: I52905
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 342-352,'C',354-359 <WI2>
A:Cross-references: GB:S64717; NID:g408195; PID:g408196
A:Note: mutant sequence from patient with osteogenesis imperfecta
R:Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A:Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalph
A:Reference number: A90476; MUID:84080385
A:Accession: A90476
A:Molecule type: mRNA
A:Residues: 425-1250,'X',1252-1328,'S',1330-1390,'X',1392-1464 <BER>
A:Cross-references: GB:K01228; NID:g180391; PID:g180392
A:Note: sequence partially completed for missing nucleotides by A29439

R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multiexon deletion in an osteogenesis imperfecta variant with increased type II
A;Reference number: A22161; MUID:85104934
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594,'R',596-607 <CH3>
A;Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PID:g179615
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained

A;Reference number: A35336; MUID:90252792
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720,'E',722-737,'E',739-745 <WAL>
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the α
A;Reference number: I54365; MUID:95187161
A;Accession: I54365
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 746-766,'S',768-781 <FOR>
A;Cross-references: GB:L47667; NID:g1009093; PID:g1009094
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A;Reference number: A47426; MUID:93352646
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1336,1339-1387,'R',1389-1464 <CHE>
A;Cross-references: GB:S64596; NID:g407589; PID:g407590
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIP:136445)
A;Note: does not represent an experimentally determined sequence but three different mut
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1464 <CH5>
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336,1339-1464 <CH6>
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH7>
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nic
J. Biol. Chem. 263, 14605-14607, 1988
A;Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide o
A;Reference number: I55269; MUID:89008319
A;Accession: I55269
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1187-1194,'C',1196-1220 <COH>
A;Cross-references: GB:M23213; NID:g340842; PID:g499622
A;Note: mutant sequence from a patient with mild osteogenesis imperfecta
R;Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.
Nucleic Acids Res. 16, 349, 1988
A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

alignment_scores:
    Quality: 343.00   Length: 952
    Ratio: 0.858   Gaps: 62
Percent Similarity: 42.017   Percent Identity: 27.941

alignment_block:

US-09-030-606-110/rev x CGHU1S

Align seg 1/1 to: CGHU1S from: 1 to: 1464

```
2353 AGAGATGGAAACCAGGTGACTGAGTTATTCAGTTCCCAAAACCCT.. 2306
     ||||||||  |||||||                       ::||||
131  ArgAspGlyIleProGlyLeuProGlyLeuProGlyProGlyProPr 147

2305 ....TCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTA...ACCCTGA 2263
         |||               |||||||||  :::            |||
147  oGlyProProGlyProProGlyLeuGlyLeuGlyAsnPheAlaProGlnLeuS 164

2262 GCCTGGGTAATCCACCTGCAGTCCCCGCATTCCAGTGCATGGAGCCCT 2213
     ||||||         ::||||           ||:||::      ::||||
164  erTyrGlyTyrAspGluLysSerThrGlyGlyIleSerValProGlyProPro 180

2212 TCTGGCCTCCCTGTATAAGTCCAGACTGAAACCCCCTTGGAGGCCTCCA 2163
     |||                                          ::||||
181  MetGly..........................................Pr 183

2162 GTCAGGCAGCCCTAGAGACTGGGGAGAGAGGACGCCCCAGCCCC 2113
     ||||||                     ||||| ||| |||||||
183  oSerGlyPro..........................ArgGlyLeuProGlyProP 193

2112 CAGCTGTGCAGCTACGGCACCTCAGCACGCACAGGGTGGCAGCAGAGCCA 2063
     ||:::||       :::;;  |||||      ||||||
193  rOGlyGlnGlyPheGlnGlyProProGlyProGlyGluPro 208

2062 CATTACTTTGGCAGCAACAGAAACTGGGCGCCAGCCCCGGCAGCCCCATGG 2013
     |||                 |||       |||  |||||||
209  GlyGluProGlyAlaSerGlyProMetGlyProArgGlyProArgGlyProProGlyPr 225

2012 GGCTAACAGGAGCGCGGGAGCTGGA........................CCCA 1984
     :::;;::||:||::                               ||||
225  oProGlyLysAsnGlyAspAspGlyGluAlaGlyLysProGlyArgProG 242

1983 GTGAGGCAGGCCCTCCA............................CCCCAATGT 1958
     :::||  |||||||                                |||
242  lyGluArgGlyProProGlnGlyAlaArgGlyLysLeuProGlyThr 258

1957 GCTCGAAGTTTCTCACGCTGAGTATTGGCCAAGTCGCTCTTGTCAAATA 1908
     ||||||        ||      :::          ::
259  AlaGly...............LeuProGlyMetLysGlyHisArgGlyPh 270

1907 CTACCTGTGTAGCAAAGTAAATGCGACCAGACCCAGGCCTGCGGCGACAGAC 1858
     ||  :::;;:||||||       :::            |||          ::
270  eSerGlyLeuAspGlyAlaLysGlyAlaAspAlaGlyProAlaGlyProLysG 287

1857 ACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGGAGCCCATAAA 1808
     ::||||||||              :::||||||||
287  lyGluProGlySerProGly............GluAsnGlyAlaProGly 299

1807 CAGGGATGGGGCCA..........CCTGGGACAGCAGGAAGGCACTATCCA 1767
     ||| ||||||||             ||||| |||||
300  Gln..MetGlyProArgGlyLeuProGlyGluArgGlyArg......Pro 313

1766 GGATGGCGAGGTCCAGGCAGCAGATGCCCCGGCCCGGAA................ 1731
     ||  :::;;:||||||          |||       ::
314  GlyAlaProGlyProAlaGlyProAlaArgGlyAsnAspGlyAlaThrGlyAl 330

1730 ......CCACCCTGGCCTCCGTGGCCTCACCCACCACCA........ 1696
     |||||| |||                 ::||||||||
330  aAlaGlyProGlyProThrGlyProAlaGlyProPheProG 347

1695 ..........CGTACGGAGCATCACAG.............GCA 1675
     :::;;::||:||::                        |||
347  lyAlaValGlyAlaLysGlyGlyAlaGluAlaGlyGlyProGlnGlyProArgGlySer 363

1674 GAGGCCCGCCAGAGCCCGGGTGGAGGTGGAGCAGGCCACTGCCTCCAGC 1625
     ||::||||||||          |||||||| ::||||
364  GlyGlyProGlnGlyValArgGly.......GluProGlyProGlyProGl 377

1624 ACCCACGTGTCCATTAGGG...AAGGGAGCTCCAGGCTTAGGG......C 1584
```

```
377  :|||::                    |||          |||   ||||||
     yProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnP   394
1583 CTGGCAGGAAGCTGGTCATCAGGCTGCTCCTCACTGCTAGCCACCTCCAGTG 1534
     |||||  |||   ::||||||||:::|||
394  rOGlyAlaLys.GlyAlaAsnGlyAlaPro.................      403
1533 TCCCCTGGTATTTGGGCAGGAACACCTGCTTCCCGGTGGTAGAGGGA       1484
     |||||| |||||||::|||:: |||||||:||||||
404  ......GlyIle..AlaGlyAlaProGlyPheProGlyAlaArgGly.      416
1483 GCCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGA    1434
     :::||||:: ||||||||   ||  ||||||
417  .......ProSerGlyPro.GlnGlyProGlyGlyProProGlyPr       429
1433 GGGCGGCTGAAGCTGTCACCACGGCCACCACTGTGGACAGGCCATGTGGCA   1384
     |||  ||||   |||  :: |||:::: |||||||
429  oLysGlyAsnSerGlyGluProGly..AlaProGlySerLysGlyAspT    445
1383 CCGGCAGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCCGAGT  1334
     ||||| |||  |||::: ::                            ||
445  hrGly.AlaLysGlyGluProGlyPro.....................Va   454
1333 GCCGAATCGCTGCACCAGCCGGTCCATGACCAGAG.....AGAAGACCA    1290
     :: ||||:: ||||                    :::
454  l.GlyValGlnGlyProProGlyProAlaGlyGluGluGlyLysArgGly    470
1289 GGGAGATGGGCACTGCAGGAACAGCC...CCAGGCTGCCCATCCGAACG    1243
     :::|||  :: :||:    |||||: |||::||||
471  AlaArgGlyGluProGlyProThrGlyLeuProGlyProGlyProGlyGluAr 487
1242 C........................CTTCATCATAGTGTCTCCCGGC     1220
     :::|||                    ::|||||:|||
487  gGlyGlyProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyP   504
1219 CTCGGTGCCCGGCTCAGCCTCTGGGCACGCCCTGGTACAGCCCCTCGCCCA  1170
     ||::: ||||||          ||||| ||||||
504  rOLysGlyProAlaGlyLysArgGlyAlaGlySerProGlyProGlyProLys 520
1169 CGAAATCCGTGTAAAACAGGCTGAAGGTCATGAGTGCCATCCAGCTGCAC   1120
     :::|||||
521  GlySerPro....................................      523
1119 AGCTCAGCCACGAAGAGCGGGCCCAGGGTGCGGGCATGCGGCAGCACAG   1070
     :::|||:::|||||| ||||::: |||
524  ......................................GlyGluA     526
1069 CTGGTGCAGCCGGGAAGCAGGGGCGCCCAGGT.........TCCGGA    1032
     |||||||||||| |||||                    ::||
526  laGlyArgProGlyGlnAlaGlyLeuProGlyAlaLysGlyLeuThrGly    542
1031 AAGCCAAGGCGCCCGGACGCCATGAGACCAGTGGGCGACGAGGAGGGGCC    982
     :::||||::::||||| |||:: ||||||
543  SerProGlySerProGlyAspGly.........LysThrGlyPr        555
981  GACAGCCCTTCTGCTGCTCGCGTGGGCCCAGCG...................  948
     ||||||||
555  oPro................GlyProAlaGlyGlnGlyAspGlyArgP    565
947  ..CTGCCTCCTCAGCCACCA................................  930
     |||:: |||||
565  rOGlyProProGlyProProGlyProGlyAlaArgGlyAlnAlaGlyValMetGly 581
929  ..........GCAGTGTGGCCTGCTACGGAGGTGAGGAAGATG        898
     |||::: |||
582  PheProGlyLysGlyAlaAla....................GlyGluPr    592
897  AGGGTGAGCAGGCCAAAGAGGCACT.....CCTCCTGGGTGCCCAGGT.    855
     |||  |||:::||||||:::       |||||
592  oGlyLysAlaGlyGluArgGlyValProGlyProProGlyAlavalGlyP   609
854  ...................................AGG             852
```

```
609  rOAlaGlyAlaGlyLysAspGlyGluAlaGlyAlaGlnGlyProProGlyProAla 625
851  GGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA..........    813
     |||||||| |||   ||||||
626  GlyProAlaGlyGlyArgGlyGluGlnGlyProAlaGlySerProGlyPh   642
812  ...........GGTAGCCCAGGCAGCCCCCAAGACTGATCATGATGAAGGC  776
     |||   |||
642  eGlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProG   659
775  ATAGACAGAGTAGGCCTGGCACAGTGGT...CCGGGTCCC.....GGA    735
     :::|||||::  |||||||| ||  ||||||
659  lyGluGlnGlyValProGlyAspLeuGlyValAlaProGlyProSerGlyAla 675
734  AGAGGTCAGAGACGAGGCCTCCAGTGAGTGAAGCACACCTGCCACCAG    685
     ||||||||||||::            ||||||
676  ArgGlyGluArg...GlyPheProGlyGluArgGlyValGlnGlyProPr  691
684  AAGTCCAGCAGCCCACGCCCAGGATGAGCAGTGCCA.............    648
     :::||||||::::|||| |||:: :::|||
691  oGlyProAlaGlyProArgGlyAlaAsnGlyAlaAlaProGlyAsnAspGlyA 708
647  ...............GCTCCAGGGGCCTGGGATCCGGGCACAGCAGCC    615
     ||||||||||||            :::
708  laLysGlyAspAlaGlyAlaProGlyAla......ProGlySerGlnGly  722
614  CTGCTAGCCCAGCCGGCCCTTGGGATGAGAAGAAGGCCTCAGCAGGATGCCC  565
     ||::|||||||    :::    ||||||||::::||||
723  AlaProGlyLeuGlnGlyMetProGlyGluArgGlyValArgGlyLeuPr   739
564  AAGGACAGTGCCCAGATGAAGGGCGGCGGCCCATAGCGTCCACGCCA     515
     :|         ||||||||   :::  ||::: :::|||||
739  oGly......ProLysGlyAspArgGlyAspAspGlyProLysGlyAlaA   754
514  GTGGTCACTGCCTGAGCCTAGGAGCG...GGACACAGA.........    480
     ::|||      |||             |||
754  spGly..........SerProGlyLysAspGlyValArgGlyLeuThrGly 767
479  ...............CCAGGCCCAGCACTGACCAATGCCA........    453
     ||||||||||||
768  ProIleGlyProProGlyProAla......GlyAlaProGlyAspLysGl  782
452  ..............................................GCAC  449
     |||
782  yGluSerGlyProSerGlyProAlaGlyProThrGlyAlaAArgGlyAlaP  799
448  CATGGTCATGAACTTCTCCTCTACCCCACTTCCAGCAGCAGAGGCGGCA    399
     ||  :::|||        ||||||::::|||
799  rOGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly   815
398  CATAGGTGATGCCTGCGGCCAAACACACCCTCCAGCC.........CA    361
     ||||||||
816  ..........................ProProGlyAlaAspGlyGlnPr   823
360  AAGGTTAGCAGGTTGACCAGCAAGAGTCTGGGCTTCCGGTGCCCAGCAG    311
     |||   |||:: |||||| |||||   ::||||
823  oGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGlyAspAla....  838
310  GCGGCTCACCCAGCCTCTGGACCATAGTGGGCCA............GG    273
     |||  |||    ||||||              ||
839  .GlyProProGlyProAlaGlyPro.AlaGlyProProGlyProIleGl   854
272  CGGGTAGGGCTCAGGGGGCCGTTCAGGCCACTCCAGCTGCTTCGTCTCG    223
     |:::  |||  |||:::|||        |||
854  yAsnValGlyAlaProGlyAlaLysGlyAla.................ArgG 866
222  GCTCTGCT......CCAGAAGCTGCGGCCTCTCCTCCTTGCTCCGCCAA    179
     |||  |||:::       |||    |||:::::||||
866  lySerAlaGlyProGlyAlaThrGlyPheProGlyAlaAlaGlyAlaArg  882
178  CTGCCTAGGAATCAGCCAGGCGCCCATTCTGCCAGCCCTTTGCTCCGGG    129
     :::||||||||::::|||
883  ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGl   899
```

```
128  TCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGG......GCAC   85
     |||||||::::                              |||
899  yProAlaGlyLysGlu..................GlyGlyLysGlyP    909

 84  CTCAGTGGGACACGTCTCATCACTCAGATCCTGGCCGA...........   46
     ||:|| |||:::|||::::         ::: ||||||::
909  roArg.GlyGluThrGlyProAlaGlyArgProGlyGluValGlyProPr  925

 45  ..............GGCGCGGCGTGTCACCCGGAGCC   22
                  |||  :::  |||||||||||
925  oGlyProProGlyProAlaGlyGlyGluLysGlySerProGlyAla   939
```

seq_name: pir1:CGCH1S

seq_documentation_block:
collagen alpha 1(I) chain - chicken (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 21-Jul-1995
C;Accession: A90458; A90181; A02857
R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete prima
A;Reference number: A90458; MUID:82231995
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Experimental source: skin
A;Note: this is the latest in a series of papers from these workers elucidating the sequ
R;Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A;Reference number: A90181; MUID:72243016
A;Accession: A90181
A;Molecule type: protein
A;Experimental source: skin
A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C;Comment: Most of the prolines at the third position of the tripeptide repeating unit (
C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

alignment_scores:
        Quality: 342.00   Length: 987
        Ratio: 0.783   Gaps: 62
Percent Similarity: 44.276   Percent Identity: 27.356

alignment_block:
US-09-030-606-110/rev x CGCH1S ..

Align seg 1/1 to: CGCH1S from: 1 to: 1042

2578 ATGGCAACAGAAGGACCAACAGGCCACATCCTGATAAAAGGTAAGAGGGG 2529
      |||::: ::||||||::|||| ||| |||||::
  55 MetGlyProArgGlyProAlaGly..ProProGlyLysAsnGlyAspAsp   70

2528 GGTGGATCAGCAAAAGACAGTCGCTGTGGCGTCGAGGGACCT........ 2487
      |||||::::::: :::|| ||||| |||||||::
  71 GlyGluAlaGlyLysProGlyArgProGlyGlnArgGlyProProGlyPr   87

2486 ....GGTTCTTGTGTGTGCCC...CTCAGGACTCTTCCCCTACAAATAA 2444
      ||::: |||||| ||||| ::: ||||
  87 oGlnGlyAlaArgGlyLeuProGlyThrAlaGlyLeuProGlyMetLysG 104

2443 GTCATATG.......................TTCAAATCCCATGGAGGAGTG 2415
      ::||| ||:|||||
 104 lyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAlaLysGlyPro 120

2414 TTTCATCCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGG 2365
```

```
121  AlaGlyProLysGlyGlyGluProGlyLysProGlyGluAsnGlyAlaProGlyI  137
     |||::::::::                      |||  |||    ||
2364  TTAAGGGGCTTAGAGATGGGAAACCAGGTGACTGAGTTTATTCAGCTCCC  2315
      :::||| ||| ||| |||   |||||:::                :::||
137  yGlnMetGlyProArg.GlyLeuProGlyGluArgGlyArgProGlyPro  153

2314  AAAAACCCTTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT  2265
      ::::::||:::            ::::: |||::
154  SerGlyProAlaGlyValAlaArgGlyAsnAspGlyAla..........  165

2264  GAGCCTGGGTAATCCACCTGCAGAGTCCCG......CATTCCAGTGCAT  2221
      ||| |||   :::|||                   :::||||
166  ..........ProGlyAlaAlaGlyProProGlyProThrGlyPproA   178

2220  GGAGCCCTTCTGGCCTCCCT.....GTATAAGTCCAGACTGAAACCCCC  2177
      :::||| |||::::||       ::: :::||  |||||
178  laGlyProProGlyPheProGlyAlaAlaGlyAlaLysGlyGluThrGly  194

2176  TTGGAAGGCCTCCAGTCAGGCAGCCCTAGAGACTGGGGAGGAGGAGAGG  2127
      ::|||  |||||   ||||::    |||:::      ||||
195  ProGlnGlyAlaArg...GlySerGluGlyProGlnGlySer...ArgGl  209

2126  GACGCCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGGGTG  2077
      |  ||| |||||::::                    |||
209  yGluProGlyProProGly.............ProAlaGlyAlaAlaGlyP  222

2076  GCAGCAGAGCCACCATTACTTTGCAGCAACAGAAACTGGGCGCCAGCC  2027
      ::: |||||||::           :::          |||
222  roAlaGlyAsnProGlyAlaAspGlyGln.ProGlyAlaLysGlyAlaTh  238

2026  CGGCAGCCCCATGGGCTAACAGGAGCGGGGAGCTGGGACCCCAGTGAGGC  1977
      ||:::|||  |||:::::||||||    :::::   :::|||
238  rGlyAlaPro...GlyIleAlaGlyAlaProGlyPheProGlyAlaArg.  253

1976  AGGGCCTCCACCCCAATGTGCTGGAAGTTTCTACGCTGAGTATTTGGCC  1927
      |||||                                        |||
254  GlyProSerGlyProGln................................GlyPr  261

1926  AAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAG  1877
      ||||  :::                                       |||
261  oSerGlyAla......................................ProG   266

1876  ACCCA...........GGCCTGCGGCAGACACCATATAGGCAGTGACAG  1839
      |||              ::|||  |||:::  :::
266  lyProLysGlyAsnSerGlyGluProGlyValAlaProGlyAsnLysGlyAsp  282

1838  ACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATGGGCGCCACCTGG  1789
      |||||| |||::::  |||||::::||                :||||||||||
283  ThrGlyAlaLysGlyGlyProGlyProAlaGlyValGlnGlyProProGlyI  299

1788  GACAGCAGGA..............AGGCACTATCCAGGATGGCGAGTCCAG  1751
      |  ||||||                   |||    |||    |||||::
299  yProAlaGlyGluGluGlyProGlyArgGlyAlaArgGlyAlaGlyProA  316

1750  GCAGATCCCCGGCCCCGGAA.....................  1731
      ::  ||||||||
316  laGlyLeuProGlyProAlaGlyGluArgGlyAlaProGlySerArgGly  332

1730  ...............................CCACCCTGGCCTCG  1717
                                     ||||||
333  PheProGlyAlaAspGlyIleAlaGlyLysGlyProProGlyGlyLyGluAr  349

1716  GTGGGCTCACCCACCACCACTACGGAGCAATCACAGGCAGAGGCC..  1669
      |  |||||||||   ::::: :::                |||||||||
349  g..GlySerProGlyAlaValGlyProLysGlySerProGlyGluAlaGl  365

1668  ....CCGCAGAGCGCGGTGGAGGTGGGAGCAGGCCACTG...........  1633
          :::||||||                         |||||||
365  yArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySerP  382

1632  ...................................CCTCCAGCACCCACCTGT  1615
                                         ||||||:::|||::
```

```
382  roGlySerProGlyProAspGlyLysThrGlyProProGlyProAlaGly  398
1614 CCATTAGGGAAGGGAGGAGCTTCAGGCTTAGGCCTGCCAGGAAGCTGGTCAT 1565
      |||||      ::|||||:  ||||||:::
399  GlnAspGlyArgProGlyProAlaGlyProProGlyProAlaArg........ 412
1564 CAGGCTGCTCCACTGCTAGCACCTCCAGTGTCCCCTCCGGTATTGGGCA 1515
      |||||:::::|||:::||   |||  |||   |||
413  .GlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAla..AlaG 427
1514 GGAACACCTGCTTCTCCGGTGGTAGAGGGAGGCCA.............. 1479
      |||  |||::::|:||||  ||||| |||
428  GlyGluProGlyLysProGlyLysArgGlyAlaProGlyProProGlyAl 444
1478 .......GTGTGTAGGGCAGGAATCTCAGGGCTGAGAAGGTGAACCC 1439
      :::||||||| ::||||  ||||:|||
444  avalGlyAlaAlaGlyLysAspGlyGluAlaGlyAla.GlnGlyProPro 460
1438 GGTGAGGGGCTGAAGCTGTCACCAGGCCACACTGTGGACAGGCATG 1389
      |||   |||   :::||||||:::||||:::     |||   |||
461  GlyProThrGlyProAlaGlyLyGluArgGlyGluGln...GlyProAla.G 476
1388 TGGCACCGGCAGCCACAGGGAAAGCTGCCACACTGCCAAATAGAGACTGCT 1339
      |||||:::  |||||||
476  lyAlaProGlyPhe.GlnGly..................... 482

1338 CGAGTGCCAATCGTCGCCAGGCTCGGTCCATGACCAGAGAGA...AGAC 1292
      |||||||::: :||||||||    |||   |||||   |||:|
483  .......LeuProGlyProAlaGlyLyProProGlyGluAlaGlyLysP 496
1291 CAGGGAGAATGGCGGCACTGCAGGAACAGCC.............. 1263
      |||||:::  |||  |||::|||
496  roGlyGlyGluProGlyValProGlyAsnAlaGlyAlaProGlyProAlaGly 512
1262 ...........CCAGGCTGCCCATCCGAACGCCTTCATCATAGTGTCTC 1225
      |||  |||::::||||||
513  AlaArgGlyGluArgGlyPhePproGlyGluArgGlyValGlnGlyProPr 529
1224 CGGGCCTCCGTGCCCGGCTCAGCTCTGGGCACGCCCTGGTACA....... 1182
      |||||||||||||| ::|||||
529  oGlyProGlnGlyProArgGlyAlaAsnGlyAlaProGlyAsnAspGlyAlyA 546
1181 ......GCCCCTGCCCACGAAATCCGTGTAAAACGCGTTGAAGG 1143
      ||| |||   :::       |||  :::
546  laLysGlyAspAlaGlyAlaProGlyAlaProGlyAsnGluGly...... 560
1142 TCATGAGTGCCATCCAGCTGCACAGCTCAGCGACGCCACGAGAGCCGGCCAGG 1093
      |||   :::      :::::|||   |||::: ||||||
561  .......ProProGlyLeuGluGlyMetProGlyGluArgGlyAla.. 573
1092 GTGCGGGCATGCGGCAGCACAGTCGGTCAGCCGGGGAAGCCAGGGCGCC 1043
      |||||:::::  |||||   |||::: |||   |||
574  ..........AlaGlyLeuProGlyAlaLysGlyAspProGlyArgGlyAspP 586
1042 CAGGT...TCCGGAAAGCCAAGGGGCCGGCATGGAGCAGCAGTGGGGGCG 996
      ||||::|| |||||||  ::|||   |||:::
586  roGlyProLysGlyAlaAspGlyAlaProGlyLysGlyAspAspGly..... 599
995  ACAAGGAGGGGCGCCGACAGCCCTTCTGCTGCTCCGTGGGCCCAGCGCT 946
      ||||||    |||  |||
600  ...LeuArgGlyLeuThr.......................GlyProIleGl 608
945  GCCTCCTCAGCCA.............................. 933
      |||:::  |||
608  yProProGlyAlaProGlyAlaProGlyAspLysGlyGluAlaGlyProP 625
932  ....CCAGCAGTGTGCTCTACCGCAGGTAGGAAGATGAGGGTGAGCA 888
      |||||    :::::|||||
625  roGlyProAla.......................GlyProThr 631
887  GGGCCAAAGAGGCACTCCT.............CCTGGGTGCCCAGGT.. 855
      |||  ||||||:::|||              |||||||:::|||
632  GlyAlaArgGlyAlaProGlyAspArgGlyLyGlyProGlyProProGlyPr 648
```

```
854  .........AGGGGGCCAGGCAGTGGCACTGGTGTCCCAGTCAATGGCAGGCAG 815
          |||||:::
648  oAlaGlyPheAlaGlyPro..........................ProGlyAlaAspGlyGlnP 661
814  GAGGTAGCCCAGGCAGCCCCAAGACTGATCATGAAGGCATAGACAGAGT 765
          :::||||  |||   |||
661  roGlyAlaLysGlyGluThrGlyAspAlaGlyAlaLysGlyAspAlaGly 677
764  AGGCCTGGCGACAGTGGTCCG........................... 744
      |||||   |||   |||
678  ProProGlyProAlaGlyProThrGlyAlaProGlyAlaGly***Va 694
743  ....GGTCCCGA...AGAGGTCAGGA......GCAGGGCCTCCAGTGG 707
      :::|||||| ::|||||   |||      |||||||||
694  lGlyAlaProGlyProLysGlyAlaArgGlySerAlaGlyProProGlyA 711
706  AGTGAAGCACACCT.......................GGCCACAGAAGTCCAGCAGCCCCACGC 666
      |||||||:::||           :::|||
711  laThrGlyPheProGlyAlaAlaGlyLyArgValGlyProProGlyProSer 727
665  CCAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATCCGGCACAGCAGC 616
      :::  |||||::::|:|||     |||||    |||
728  GlyAsnIleGlyLeuProGlyPro..............ProGlyProAlaGl 740
615  CCTGCTAGCCAGCCGG..CCCTTGGGATGAGAAAGAGGCTCAGCAGGAT 569
      ::::||||: |||       |||   |||::: |||||
740  yLys***GlySerLysGlyProArgGly...GluThrGlyProAlaGlyA 756
568  GCCCA.......................AGGACAGTGCCCAGATGAAGGGCCGGCGGCGC 531
      |||                         :::|||
756  rgProGlyGluProGlyProAlaGlyProProGlyProGlyGluLys 772
530  CATAGGCTCCACGCACCAGTGGTCACTGGCTGCTGAGCCGGACACAG 481
      ::|      |||::::|||     ||||||
773  GlySerProGlyAlaAspGly.........ProIleGlyAlaProGlyTh 786
480  ACCAGGCCA.............................. 471
      ||||||||
786  rProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGlyLeuProG 803
470  ................................GCACTGGACCAATGCCCAGCCAGCACCATGGTCATGA 438
      :::|||:::|||
803  lyGlnArgGlyLeuArgGlyPheProGlyLeuProGlyProSerGly... 818
437  ACTTCTCCTCACCCCCACTTCCAGCACAGCAGAGGCGGCCACATAGGTGATG 388
      :::|||      |||   |||::::|||
819  ...GluProGlyLysGlnGlyProSerGlyAlaSerGlyAlaArgGlyPr 834
387  CCTGCGCGCCAAACACACCTCCAGGCCCAAAGGTTAGCACAGTTGACCAGCAA 338
      |||||||   ||||||    :::|||
834  oProGlyProMetGlyProProGly.......LeuAlaGlyProProGlyG 849
337  GAGCTGGGCTTTCCGGTGCC...GCAGCAGGCGGCTCACCCACCAGCCTCT 291
      |||||||:::|||      ::|||:::|||
849  luAlaGlyArgGlyValAlaProGlyAlaGluGlyValAlaProGlyArgAsp 865
290  GGACCATAGTGGGCCA...GCCGGGTAGGGCTCAGGGGGCCGTTCAGGCA 244
      ||:||||     :::|||   :::||||||
866  GlyAla.AlaGlyProLysGlyAspLysGlyGluAlaGlyAlaGlyP 882
243  CTCCAGAACTGCTTCGTCTCCGGCTCCTGCTCCAGAAGCTGCGGCCCTCCT 194
      |||         :::|||      ||||||:::|||
882  roPro.........................GlyAlaProGlyAlaProGlyAlaPro 892
193  ................CCTTGCTGCCGCCCAACTGCCTAGGAATCAGCCAGGCGC 156
      |||                  :::|||   |||
893  GlyProValGlyProAlaGlyLysAsnGlyAspArgGlyGluThrGlyPr 909
155  CCATTTCTGCCAGCCCTTTGGTCCGGTCCAGCTTCTCAGCCC 113
      |||||||:::|||              :::|||
909  o.......................AlaGlyProAlaGlyProPro 916
```

Wed Sep 29 14:26:56 1999

seq_name: pir2:S18803

seq_documentation_block:
collagen alpha 1(V) chain - hamster
C;Species: Cricetinae gen. sp. (hamster)
C;Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 16-Dec-1998
C;Accession: S18803
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
J. Biol. Chem. 266, 24727-24733 1991
A;Title: The pro-alpha(V) collagen chain. Complete primary structure, distribution of e
A;Reference number: S18802; MUID:92105142
A;Accession: S18803
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1843 <GRE>
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
F;1620-1842/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

alignment_scores:
        Quality: 335.50   Length: 1216
          Ratio: 0.693   Gaps: 78
Percent Similarity: 39.803   Percent Identity: 24.836

alignment_block:
US-09-030-606-110/rev x S18803

Align seq 1/1 to: S18803   from: 1 to: 1843

3068 GGTCCTCGAGTAGCTCCAAACAGGGTTGTGGAGCTGGTGGGAAAGTTGG 3019
     |||||||||::::| :: :| ||| ||
 436 GlyProGlyMetProAlaAsnGlnAspThrIleTyrGluGlyLyIleGlyGl 452

3018 GGGTAGGGGAAAGTTGGGG...GTAGGGGAAATTTGGCAGTGCCTTCA 2972
     | |||||||::: ||| |||:||||::|||||
 452 yProArgGlyGluLysGlyGlnLysGlyGlu.........ProAlaIleI 466

2971 TCAGCCCAGTCCTAGAGAGAGGAGAGGG.................. 2943
     ||:::||| :: :|||||
 466 leGluProGlyMet...LeuIleGluGlyProProGlyProGluGlyPro 481

2942 ..................AGTGGAAGTGGGGGAACCAGGC 2920
                      ::||| |||||||::|:
 482 AlaGlyLeuProGlyProProGlyThrThrGlyProThrGlyGlnValGl 498

2919 TGGGCCAAGAGAAGAGAGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCC. 2871
     ||| |||||||| ||||||| |||
 498 yAspProGlyGluArgGlyGlyProProGlyArgProGlyLeuProGlyAlaA 515

2870 ..........CCACCCTCTACCTTCCTTCAACACCCTAACCTTGGG 2835
               |||||||||:||::||||| ||
 515 spGlyLeuProGlyProGlyThrMetLeuMetLeuPro.......... 528

2834 TAACAGCATTTGGAATTATCATTTGGATGAGTAGAATTTCCAAGGTCCT 2785
        || ||||||| ||| ||| ||||||
 529 ....PheArgPheGly................ 532

2784 GGGTTAGGCATTTGGGGGGCCAGA...............CCCCAGGAGA 2750
     ||| :: ||||||::   :|||||:
 533 .GlyGlyGlyAspAlaGlySerLysGlyProAsnValSerAlaGlnGluS 549

2749 AGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGA... 2703
     ::||:|||||||||| ||||||||||
 549 erGlnAlaIleLeuGlnGlnLeuGlnAlaArgLeuAlaLeuArgGlyPro 565

2702 ...................CCTGATTGTTGGGGATCCCCCAC 2680
                        ||| |||||
 566 AlaGlyProMetGlyLeuThrGlyArgProGlyThrMetGly.ProProG 582

2679 CCTACCCAAATATTAGAACCACCAACAAGAAAAGCTAGCAATGGATTCCCT 2630
       :::::||||| : ::   :: ||
 582 lySerGlyGlyLeuLysGlyGluProGlyGluProGlyAspMetGlyProGln 598

2629 TCTACTTTGTTAAATAAATAAGTAAATATTAAATGCCTGTGTCTCTGT 2580
                          :::||
 599 Arg...................................GlyVa 601

2579 GATGGCAACAGAAGGACCACCAGGCCACATCCTGATAAAAGGTAAGAGG 2530
     |||:::|| :::|||:::|
 601 lGlnGlyProProGlyProThrGlyLysProGlyArgArgGlyArgGlyAlaG 618

2529 GGGTGGATCAGCAAAAGACAGTGCTGTGGGCTGAGGGGACCTGGTTCTT 2480
     || ||| :::
 618 lySerAsp............................. 620

2479 GTGTGTTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAA 2430
     |||||||||
 620 620

2429 TCCCATGGAGGAGTGTTTCATCCTAGAAACTCCCATGCAAGAGCTACATT 2380
     |||||||:||| |||||::||
 620 620

2379 AAACGAAGCTGCAGGTTAAGGGGCTTAGAGATGGAAACCAGGTGACTGA 2330
     ||||||| ||| ||| |||
 621 ....GlyAlaArg...GlyMetProGlyGln... 628

2329 GTTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTCTTCAACTAGGAGG 2280
     ::|||||||||:| :: ||| :|| |||::|||||
 629 ......ThrGlyProLysGly...AspArgGlyPheAspGlyLeuAlaGl 642

2279 CTAGCTGTTAACCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATT 2230
     |||:::|||:|| |||||||| |||
 642 Y..................LeuProGlyGluLysGlyHisA 650

2229 CCAGTGCATGGAGCCCCTTCTGCCTCCCTGTATAAGTCCAGACTGAAACC 2180
     ::||||||||| :|||||
 650 rgGlyAspGlyProSerGlyProPro................ 659

2179 CCCTTGGAAGGCCTCCAGTCAGGCAGCCTAGAGACTGGGAGAGAGG.. 2132
     ||| ||| ||||||| ||||||||||
 660 ....GlyLeuProGlyAspAspGlyGlyArgGl 669

2131 ....................AGAGGGACGACCCCCAGCCCCCAGCTGTGC 2104
                         ||||| |||||::: :
 669 yAspAspGlyGluValGlyProArgGlyLeuProGlyGluProGlyProA 686

2103 AG..........CTACGGCACCTCAGCACCACAGGGTGCCAGCAGAGCCCACAT 2060
       ||| :: ::|| : |||::|||||
 686 rgGlyLeuLeuGlyProLysGlyProProGlyProGlyProPro... 701

2059 TACTTTGGCAGCAACAGAAACTGGGCGGCAGCCTCAGCCCCATGGGGC 2010
     ||| ::||||||||||||||||||||| ||||||
 702 ....Gly.ValThrGlymetGlyMetAspGlyGlnProGly...ProLysGlyA 715

2009 TAACAGGA...GCGGGGAGCTGGGACCCAGTGAGGCGGCCCCTCACCCC 1963
        :|| |||||||| |||
 715 snValGlyProGlnGlyGluProGlyGlnGlnGlyGlnGlyGlyAsnPro 731

1962 AATGTGCTGGAAGTTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTC 1913
     ::: |||::| ||| |||:::
 732 Gly.............AlaGlnGlyLeu.ProGlyProGlnGlyA 742

1912 AAATACTACCTGTGTAGCAAAGTAAATGGCAAATGGCGGACCCAGGCCTGCGG 1863
     ||| ||| |||||||||
 742 laIleGlyProProGlyGluLysGlyProLeuGlyLysProGlyLeuPro 758

1862 CAGACACCATATAGGCAGTGACGACTGGCTGAGCTGGAACAATGGAGCCC 1813
     ||: || :::
 759 GlyMetProGlyAlaAspGlyProProGlyHisProGlyLys....... 772

1812 ATAAACAGGATGGGCCGGCCACCTGGG...ACAGCAGGAAGGCACTATCCAG 1766
     ||||||||| ::::|||| ::: ||||
 773 ....GluGlyProProGlyGluLysGlyGlyLysGlyGlnGlyProProG 786

1765 GATGGCGAGGTCCAGGCAGATGCCCGGCCCCGGAA.......... 1731

```
1014 lyProMetGlyGluArgGlyHisProGlyProProGlyProProGlyGlu 1030
1019 .........CCCGGCATGGACAGCAGCAGTGGGCACCAAGGAGGAGGGGCCGAC 979
1031 GlnGlyLeuProGlyValAlaGlyLyLysGluGlyThrLysGlyAspPro. 1046
978 AGCCCTTCTGCTGGCTCGGTGGGGCCCAGG...CTGCCCTCCTCAG.... 936
1047 .................GlyProAlaGlyLeuProGlyLysAspG 1056
935 ..CACCCAGCGTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCA 888
1056 lyProPro..................GlyLeuArgGlyPhePro 1064
887 GGCCAAAGAGGCACT.................................. 873
1065 GlyAspArgGlyLeuProGlyProValGlyAlaLeuGlyLeuLysGlyse 1081
872 ......CCTCCTGGGTGCCCAGGT....AGGGGCCCAGGGCACTGGT 836
1081 rGluGlyProGlyProProGlyProAlaGlySerProGlyGluArgG 1098
835 GTCCCAGTCAATGGCAGGCAGGAGGTAGCCCAGGCAGCCCCAAGACTGA 786
1098 lyProAla......GlyAlaAlaGlyProIleGlyIleLysPro...... 1109
785 TCATGAAGGCATAGACAGAGTAGGCCTGGCGACAGTGGTCCGGGTCCCGG 736
1110 ..............GlyArgProGlyProGlnGlyPro........ 1117
735 AAGAGGTCAGAGAGCAGGGCCTCCAGTGAGTGAAGCACACCTGCCACA 686
1118 ...........ProGlyProAlaGlyGluLysGlyValProGlyGluL 1130
685 GAAGTCCAGCAGCCCCCAGGATGAGCAGTGCCAGCTCCAGGGGCC 636
1130 ysGlyProGlnGlyPro.AlaGlyArgAsp.....GlyLeuGlnGlyPr 1144
635 TGGGATCCGGGCACAGCAGCCCTGTAGCCACCCGCCCCTGGATGAG. 587
1144 o...ValGlyLeuProGlyProAlaGlyProValGlyProGlyGluA 1160
586 ..............AAAGAGGCTCAGCAGGATGCCCAAGGA 560
1160 spGlyAspLysGlyIleIleGlyGluProGlyGlnLysGlySerLysGly 1176
559 CAGTGCCAGATGAAGGGCCGGCGGGCCCATAGGCTCCACGCCAGTGGT 510
1177 AspLysGlyGluGlnGlyLyProProGlyPro..GlnGlyP 1192
509 CACTGCTGAGCCT...AGGAGCCGGGACACAGACCAGGCCCAGCACTGA 463
1192 roIleGlyGlnProGlyProSerGlyAlaAspGlyGluPro......Gly 1206
462 CCAATCCCCAGCACCATGGTCATGAACTTCTCCTCACCCCACTTCCAG 413
1207 ProArgGlyGlnGlyGlyLeu.......PheGlyGlnLysGlyAspGluGl 1221
412 CAGCAGGGCGGCACATAGGTGATGCCTGCGGCCAAACACCTCCAGGC 363
1221 ySerArgGly............PheProGly.ProProGlyProValGly 1233
362 ...CAAAGGTTAGCAGTTGACCACCAAGAGCTGGGCTTTCCGGT..... 321
1234 LeuGlnGlyLeuProGlyProProGlyGlyLyGluThrGlyAspVa 1250
320 ..........GCCCAGCAGGGCGGCT 305
1250 lGlyGlnMetGlyProProGlyProProGlyProArgGlyProSerGlyA 1267
304 CACCCACAGCCTCTGACCATAGTGGCCA.................. 275
1267 laProGlyAlaAspProGly.GlnGlyProGlyProGlyIleLeuGlyAsnPr 1283
```

```
786 lyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 802
1731 .................................................. 1731
803 AspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspGl 819
1731 .................................................. 1731
819 yPheProGlyPheLysGlyAspMetGlyIleLeuLysGlyAspArgGlyGluI 836
1730 .....CCACCCTGGCTCGGTGGCTCACCCACCACACGTACGGAGA 1686
836 leGlyLeuProProGlyProArg..GlyGlu................. 844
1685 CATCACAGGCAGCAGAGCCCCGAGAGCGCCGGTGGAGGTGGGAGACAGGCCA 1636
845 ..AspGlyProGluGlyProLysGlyProGlyArgGlyGlyProAsnGlyAspPro 860
1635 CTGCCTCCAGCACCCAGTGTCCATTAGGGAAG...GGAGCTCCAGGCTT 1589
861 GlyProLeuGlyProThrGlyGluGlyLyLysGlyLeuGlyValProGlyLe 877
1588 AGGG.......CCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAG 1545
877 uProGlyTyrProGlyArgGln......................G 885
1544 CACCTCCAGTGTCCCCTCGGTATT.....TGGGCAGGAACACCTGCTTC 1501
885 lyProLysGlySer.IleGlyPheProGlyPheProGlyAlaAsnGlyIGl 901
1500 TCCCGGTGGTAGAGGGAGGCCAGTGTGTAGGGCAGGATCCGAGGGCTGA 1451
901 uLysGlyYArgGlyThrPro...........GlyLysProGly..... 912
1450 GAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCACCACGGCCACACTGT 1401
912 .................................................. 912
1400 GGGACAGGCGCATGTGCACCGCCAGGCAGGGAAAGCTGCCCACACTGGCC 1351
913 ....................................ProArgGlyGln. 916
1350 AAATAGACTGCTCGAGTGCCGAATCGCTGCACCGGTCCATGACCAG 1301
917 ................................ArgGlyProThrGlyProArgGlyG 925
1300 AGAGA.......AGACCAGGG...AGATGGCGCACT 1275
925 luArgGlyProArgGlyIleThrGlyLyLysProGlyProGlyGlyAsnSer 941
1274 GCAGGAACAACGCCCCA...GGCTGCCCATCCGAACGCCTTCATCATAGTGT 1228
942 GlyGlyAspGlyProAlaGlyProProGlyProGlyGluArg........ 953
1227 CTCCGGGCCTCGGTCCCGGCTCAGCTCTGGCACGCCCTGGTACAGCCC 1178
954 ....GlyProAsnGlyProGlnGlyProThrGlyPheProGlyProLysG 969
1177 CTCGCCCCACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCA... 1131
969 lyProProGlyProProGlyLysAspGly........LeuProGly 981
1130 .................TCCAGTGCACAGCTCAGCC 1111
982 HisProGlyGlnArgGlyGlyGluThrGlyPheGlnGlyPheLysThrGlyProPr 998
1110 ACGAGAGGCGCGCCAGGGTGCGGGGCATGCGGCCAGCACA.....GCTG 1067
998 oGlyProProGlyVal...ValGlyProGlnGlyProThrGlyGluThrG 1014
1066 GTCCAGCGCGGGGAAGCAGCGGGCGCCCAGGTTCCGGAAAGCCCAAGCGGG.... 1020
```

```
  86 GlyAspArgGlyGln.....ThrGlyThrProGly.GluProGlyGlyGluL 100
2905 AGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCCCAAGCCCACCCTCTACCTTC 2856
     ::|||  |||   |||  ::: |||  |||:: ||||:::  ||||||  |||||
 100 ysGlyGluArgGlyProPro.GlyProValGlyProGlnGlyLeuProGl 116

2855 CTTCAACACCCTACCTTGGGTAACAGCATTTGGAATTATCATTTGGGAT 2806
     ||   |||                      :::          |||||||
 116 y........ValAlaGly.............HisProGly. 122

2805 GAGTAGAATTCCAAGGTCCTGGGTTAGGCATTTGGGGGCCAGACCCC 2756
     |||||| |||                  ||| |||  |||||
 123 ..ValGluGlyProGlu..............GlyProProGlyPro 132

2755 AGGAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGG 2706
     ||||||                  :: :::       ||
 133 AlaGlyArgArgGlyGluLysGlyGluProGlyArgPro........... 145

2705 GGACCTGATTGTTGGGGATCCCCACCCTACCCA..AATATTAGACACC 2659
     ||||||||     ||| :::          |||
 146 ..........GlyAspProAlaValGlyProGlyGlyAlaGlyAlaL 158

2658 AACACAGAAAAGCTAGCAAATGGAATTCCCTTCTACTTTGTTAAATAAATAA 2609
     ::  ::|||||:::::  |||:::|||
 158 ysGlyGluLysGlyAspAlaGlyLeuPro................... 167

2608 GTTAAATATTAAAATCCTGGTCGTCTCTGTGATGGCAACAGAAGGACCAAC 2559
     ||||||
 168 ...................................GlyProAr 170

2558 AGGCCACATCCTGATAAAAGGTAAGAGGGGGGTGGATCAGCAAAAAGACA 2509
     ||||||||||| ::||||||::::||
 170 gGlyAlaAlaGlyIleLysGlyGluGlnGly.................. 180

2508 GTGCTGTGGGCTGAGGGGACCTGGTTCTTGTTGTGTGCCCTCAGGACTC 2459
     |||   ||   ||||||
 181 .................................ProProGlyLeu 184

2458 TTCCCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGTGTTCAT 2409
     ::::::  ||:::::  ||:::|||
 185 ...................AlaLeuProGlyAspProGlyProLys..... 193

2408 CCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGTTAAGG 2359
     |||   |||||||                   ::::||
 193 ..............................................ArgGly 200

2358 GGCTTAGAGATGGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAAC 2309
     |||  |||||||                      ||||||
 194 ..................................GlyAspProGlyAsp. 194

2308 CCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCTGAGCCT 2259
     |||  ::||||||                 ::::||
 201 Pro...IleGlyLeuThrGlyArgAlaGly.........ProThrGlyyAs 213

2258 GGGTAATCCACCTGCAGAGTCCCCCGCCATTCCAGTGCATGGAGCCCTTCTG 2209
     :::||||||||::::||||::          :::  :||||
 213 pSerGlyProProGlyGluLysGlyAspProGlyArgProGlyProProG 230

2208 GCCTCCCTGTATAAGTCCAG.........................ACTGAAACCCCC 2177
     ||| ||||||||::|||
 230 ly...ProValGlySerGlyArgGlyArgAspGlyLeuValGlyGluLysGly 245

2176 TTGGAAGGCCTC....CAGTCAGGCAGCCCTAGAGACTGGGGAGAG 2133
     ::|||||          |||     ||| |||  ||||||||
 246 ValGluGlyAsnProGlyLeuProGlyLysAlaGlyLysGlyLeuAr 262

2132 G.....AGAGGGACGGCCC.....CAGCCCCCAGCCTGTGCAGCTACGCA 2095
     |||||||:::|||        |||||||
 262 gGlyLeuArgGlyAlaProGlyValArgGlyProAla.......... 274

2094 CCTCAGCAGCCACAGGTGGCCAGCAGAGAGCCACATTACTTTGGCAGCAC 2045
     ::  :::|||  :: ::|||           ||| |||
 275 .....GlyGluLysGlyAspGlnGlyAspProGlyGluAspGlyArgAsn 289
```

```
 274 .........GGCGGGTAGGCTCAGGGGGCCGTTCAGGGACTCCAGAAC 235
     |||                  |||:::||||  ||| |||||||
1283 oGlyAlaValAlaGlyGlulysGlyGlyProLysGlyGluAlaGlyGluProGlyL 1300

 234 TGCTTCGTCTCGGC......................................... 221
     ||    |||
1300 euProGlyGluGlyLysGlyProLeuGlyProLysGlyGluArgGlyGluLys 1316

 220 ....TCTGCTCCAGAAGCTGCGGCCCTCCTCCCTTCGTGCCGC...... 182
     ::::::|||:::::|||||||::|||||||  :::
1317 GlyGluAlaGlyProSerGlyAlaAlaGlyAlaAlaGlyProProGlyLysGlyPr 1333

 181 .......CAACTGCCTAGGAATCAGCCAGGCGCCCATTCTGCCAGC 142
     :: ::::||||||||::
1333 oProGlyAspAspGlyProLysGlyGlySerProGly............ 145

 141 CCTTTGGT.........GCCGTCCAGCTTCTCAGCCCATGCTCAAC 104
     ||||||||       :::  ||||  |||:::||
1345 rovalGlyPheProGlyAspProProGlyProGlyGluPro......Gly 1359

 103 ACCTCTGCTGTGGGGCCACCTCAGTGGGACACCGTCTCATCACTCACGATC 54
     |||||||::::       |||||| :::       ::::::|||
1360 .ProAlaGlyGlnAspGlyProGlyProGlyAspLysGlyAspAspAspGlyGluP 1376

  53 CTGGCCGAGGCGCGCGGCGCTGTCACCCGGA 25
     ||||::::: :::      |||||||||
1376 roGlyGlnThrGly.....SerProGly 1383
```

alignment_block:
US-09-030-606-110/rev x I48103

Align seg 1/1 to: I48103 from: 1 to: 1549

```
3039 GGAGCTGGTGGGGAAAGTTGGGGTAGGGAAAGTTGGGGTAGGGAAA 2990
     ||:::  |||              ::
  36 GlySerProGlyProGlnGlyProAlaGlyArgAlaGlyGluLysGlyGlyVal 52

2989 TTTTGGGCAGTGCCTTCAT.................CAGCCC.... 2965
     |||::||||                          ||||||
  52 uLysGlyAspCysGluAspProGlyAlaProGlyLeuProGlyGlnProGlyA 69

2964 .....................................AGTCCTAGA 2956
     ::|||::||
  69 laProGlyGluProGlyLeuArgGlyThrProGlyIleThrGlyProLys 85

2955 GAGAGTAGAGGGGAGTGGAGTGGGGGAACCAGGCTGGGCCAAGAGAAG 2906
     ::  ||||||| ||| |||||||
  85 GlyGluLysGlyAspGlnGlyAspProGlyGluAspGlyArgAsn
```

```
2044  AGAAACTGGGCGGCCAGCCCGGCACCCCATGGGCTAACAGGAGCGGGGA  1995
           ::: ||||:  ||| :::  ||||||
 290  GlySerProGlyProSer..GlyProLysGlyAspArg......GlyGl  303

1994  GCTGGGACCCAGTGAGGCAGGCCCTCCACCCCAATGTCTGGAAGTTTC   1945
           |||  ||||||||  ::|||:  ||
 303  uProGlyPro.....ProGlyProProGlyArgLeuValGlyAlaGlyI  318

1944  TACGCTGAGTATTTGGCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGC  1895
           ::
 318  leGly.......................................Ser   320

1894  AAAGTAAATGGCGACCA........................GA  1876
           ::: ::|||:::                              ||
 321  ArgAspLysGlyGluProGlyGlnGluGlyProProGlyArgGlyProLysGlyAs  337

1875  CCCA....................................GGCTGCGGC  1862
           |||                                       |||||||
 337  pProGlyProProGlyAlaSerGlyLeuArgGlyIleGluGlyLeuArgG  354

1861  AGACACCATATAGGCAGCAGTGACAGACTGGCTGGACAATGGGAGCCCA  1812
           |||          :::   |||
 354  lyProProGlyGlnGlyAspProGlyGlyValArgGlyProAlaGlyAsp  370

1811  TAAACAGGATGGGGCCACCTGGACAGCAGGAAGGCCACTATCCAGGATG  1762
           ||||||||||    |||:::    ||||||
 371  LysGlyAspArgGlyProProGlyLeuAspGlyArgAsn......GlyVa  385

1761  GCGAGGTCCAGGCAGATGCCCGGCCCGGAACCACCCTGCCTCCGTGGG   1712
           |||:::||                                    :::
 385  lAspGlyLysProGlyValAProGlyPro.................    394

1711  CTCACCACCACCACGTACGGAGACATCACAGGCAGAGCCCCCCAGA   1662
           |||     |||:::::|||            |||
 394  .....ProGlyProHis..GlyAlaSerGlyLysAlaGlyAspProGlyA  409

1661  GCGCGGGTGGAGGTGGGAGCAGG........CCACTG...CCTCCA   1627
           ||||||| |||     |||          :::        |||||
 409  rgaAspGlyLeuProGlyLeuArgGlyGluHisGlyProAlaGlyProPro  425

1626  GCACCCACGTGTCATTAGGGAAG.......GGAGCTCCAGGCTT   1589
           :::|||     :::|||||||:::|||         ||  :::||
 426  GlyProProGlyValProGlyLysThrGlyGluAspGlyLysProGlyLe  442

1588  AGGG................................CCTG   1581
           |:::
 442  uAsnGlyLysLysAsnGlyGluProGlyGluAspProGlyGlyArgLysG  459

1580  GCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAGCACCTCCAGTGTCC   1531
           |!::||| :::::||||||::|||                :: :::
 459  lyGluLys..GlyAspSerGlyAlaProGlyArgGluGlyProAspGlyPr  475

1530  CCTCGGTATTTGGGCAGGAGCACCTGCTTCTCCCGTGGTAGAGGGAGGC  1481
           |  |||  :::||||:::|||||||||||||||:::||   ::|||
 475  oLysGly..GluArgGlyAlaProGlyLeuArgGlyArgGlyProP   491

1480  CA.............GTGTGTAGGGCAGGAGATCTGCAGGCGGCTGAGAAGGTGA  1443
           ||            |||||||  |||
 491  roGlyLeuProGlyGln..ValGlyProProGlyInGlyPheProGlyVa  507

1442  ACCCGGTGAGGGCGGCTGAACTGTCACCACGGCCACACTGTGGGACA..  1395
           ||||||:::  |||  ::|||:::|||  :::|||||  :::|||
 507  lProGlyAsnThrGlyProLysGlyAspArgGlyGlyGluThr..GlySerLy  523

1394  ..................GGCATGTGGCACCCGGCAGCCAC  1373
                           |||  :::|||
 523  sGlyGlnGlyLeuProGlyLeuArgGlyGlyLeu.....ArgGlyGluP  538

1372  AGGGAAAGCTGCCCACTGCCCAAATAGACTTGCTCGAG.......TGC  1332
           :::||:|||||||:::    ::::|||||||||         |||  :::
 538  roGlySerLeuProAsn...AlaGluArgLeuLeuGluThrAlaGlyIle  553
```

```
1331  CGAATCGCTGCA.................................CCAGCCGG  1312
           :::||:::||                              |||
 554  LysValSerAlaLeuArgAspIleValAlaGluThrTrpGluSerSerGl  570

1311  T........CCATGACCAGAGAGAAGACCAGGAGATGGCGACTGCAG  1271
           ||   ||||||||||||||
 570  ySerPheLeuProValProGluArgArgProGlyArgProLysGlyAspProG  587

1270  GAACA.........................GCCCCAGGCTGCCCATCCGAA  1245
           ::    ||||||||||:::||:::||||||||
 587  lyGluArgGlyProProGlyLysGlyLySerIleGlyPheProGlyGlyGlu  603

1244  CGGCCTTCATCATAGTGTCTCCGGTCCTCGGTGCCGGCTCAGCTCTGGGC  1195
           |||                            ::||||||||||||||||
 604  Arg........GlyLeuLysGlyAspAProGlyAsProGlyGlnProGlyInGl  617

1194  ACGCCCTGGTACAGCCCCTGCCCACGAAATCCGTGAAACAGCGTGAA.........  1145
           : ||||| |||                  |||:::|||
 617  yProProGlyLeuAlaLeuGlyLeuGluGlyLyProProGly.........  630

1144  GGTCATGAGTGCCCATCCAGCTGCACAGCTCAGCCACGAAGAGCCGGCGCA  1095
           |||          :::|||||||||||||||:::
 631  .........ProProGlyLeuAlaGlyLyGluProGlyLysProGlyIle  643

1094  GGGTGGCGGGCCATGCGCAGCACAGCTGGTGCAGCCCGGGAAGCAGGGCG  1045
           |||      ||||||||||:::||||||||||||||||||||||||||
 644  Pro...GlyLeuProGly..ArgAlaGlySerAlaGlyValAlaGlyAr  658

1044  CCCAGGTTCCGGAAAGCCAAGGCGGGCCCGGCATGAGCAGCAGTGGGGCGA  995
           ||||||:::::  |||            :::     |||
 658  gProGlyGlyArgGlyLGluArgGlyGlyGluLysGlyAspAArgGlyGluGlnG  675

994  CAAGGAGGGGGCCGACACCCCTTCTGCTGGCGGTGGGGGCCCAGCGGCTG  945
           ||||   ||| :::     |||||||
 675  lyArgAspGlyLeuProGlyLeuPro.............GlyProProGly  687

944  CCTCCCAGCCACCAGCAGTG.................TGGCTGCTACG  913
           |||||||  |||                       || :::
 688  ProProGlyProLysValAlaIleAspGluGlnGlyProGlyLeuSerAr  704

912  CAGGTGAGGAAGATGAGGGTGAGCAGCCAAAGAGGCACTCCT......  870
           |:::   |||           |||   :::|||   |||
 704  gGluGlnGlyProProGlyLeuLysGlyAlaLysGlyLysGlyProLysSerA  721

869  ...............................CCTGGTGCCCAGGTAGGGGGCCA  846
                                       ||| ||||||
 721  spGlyAspHisGlyProLysGlyAspLysGlyAlaProGlyIleLeuLysGly  737

845  GGGGGTCCCGGTGTCCCAGTCAATGGCAGCAGGAGTAGCCCAGGCAGCCC  796
           ||||||:::|||||||:::      |||||||||||||||
 737  .GlyInGlyProProGlyLeuLysGlyGlyAlaProGlyIleLeuLysGly  754

845  GGGGGTCCCGGTGTCCCAGTCAATGGCAGCAGGAGTAGCCCAGGCAGCCC  796

738  AspGlnGlyLeuProGlyLysArgGlyHisAspGlySerProGlyLeuPr  754

795  CCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCCAGTGGT...  747
           ||::: :::                :::||||||||||::::|||
 754  oGlyGluArgGlyValAlaGlyLyProGluGlyLysProGlyLeuGlnGlyP  771

746  ...............CCGGGTCCCGGAAGAGGTCAG...AGAGCAGGGGCCTCCA  711
                         |||||||      ||| |||||||||||
 771  roArgGlyThrProGlyAlaGlyHisGlyAspProGlyProPro  787

710  GTGGAGTGAAGCACACCTGGCCACAGAAGTCAGCAGCCCACGC......  666
           ::::||||||  ||||||  :::|||||||
 788  .........GlyAlaProGlyLeuAlaGlyProAlaGlyProInGlyPr  801

665  ..................CCAGGATGAGCAG  653
                          :::||||
 801  oSerGlyLeuLysGlyGluProGlyGluProGlyGlyArgGlyL  818

652  TGCCAGCTCCA...GGGGCC................TGGGATCCGGCCACAGCA  618
           :::||::||    ||||||                 ||||||  ::
 818  euProGlyProThrGlyAlaValGlyLeuProGlyProProGlyProSer  834

617  GCCCTGCTAGCCAGCCAGCCGCGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATG  568
```

```
C:Accession: I50694
R:Nah, H.D.; Niu, Z.; Adams, S.L.
J:Biol. Chem. 269, 16443-16448, 1994
A:Title: An alternative transcript of the chick type III collagen gene that does not
A:Reference number: A54041; MUID:94266842
A:Accession: I50694
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-886 <NAH>
A:Cross-references: EMBL:U07973; NID:g520454; PID:g537432
C:Genetics:
A:Gene: COL3A1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo
F:30-90/Domain: von Willebrand factor type C repeat homology <VWH>

alignment_scores:
    Quality:   335.50               Length:   998
    Ratio:     0.826                Gaps:     65
    Percent Similarity: 40.681      Percent Identity: 27.054

alignment_block:
US-09-030-606-110/rev x I50694

Align seg 1/1  to: I50694  from: 1  to: 886
```

```
2479 GTGTGT....TGCCCCTCAGGACTCTTCCCCTACAAAATAAGTCATATGTTC 2433
          ::|||| |||   |||||   ::|||||
  51 IleCysValCysAspSerGlySerValLeuCysAspAspIleIleCysAs   67

2432 AAATCCCATNGGAGGAGTGTTTCATCCTAGAAACTCCCATGCAAGAGCTAC 2383
      :::  |||   |||||| ||||  ::||||||||| |||
  67 pAspGlnGluLeuAspCysProAsnProGluIleProLeuGlyGluCys..   83

2382 ATTAAACGAAGCTGCAGGTAAGGGCTTAGAGATGGGAAACCAGGTGAC     2333
       |||          :::|||
  84 ..........CysProValCysProGlnThrThrProGlnProThrGlu     96

2332 TGAGTTTATTCA...GCTCCCAAAAACCCTTCTCTAGGTGTGTCTCAACT 2286
      |||::||| ::|||||
  97 LeuProTyrThrGlnGlyProLys..........................  104

2285 AGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATCCA......CCTGCAG 2242
      |||::|| ||::: |||
 105 ..............................GlyAspProGlySerProGlys 112

2241 AGTCCCGCATTCCAGTGCATGAGCGCCCTTCGGCCTCCCTGTATAAGTC 2192
      ::  :::::|||:::||  :::||| |||
 112 erProGlyArgThrGlyAlaProGlyProProGlyGlnPro.........  125

2191 CAGACTGAAACCCCCTTGGAAGGCCTCCAGTCAGGCAGCCCTAGAGACTG 2142
      ||||||||||
 126 ......................................GlySerPro....  128

2141 GGGAGAGAGGAGGAGACGCGCCCCAGCCTGTGCAGCTGCAGCTACGCCACCT 2092
      |||::|||  |||||||:::::::
 129 ..........GlyAlaProGlyProProGlyIleCysGlnSerCysP      141

2091 CAGCAGCACAGGGTGGCAGCAGAGAGCCACATTAC............... 2057
        ::|||||||| :::|| :::||
 141 roSerIleSerGlyGlyLySerPheSerProGlnTyrAspSerTyrAspVal 157

2056 ...TTTGGCAGCAACAGAAACTGGCGGCCAGCCCGCCAG........... 2021
      |||::|||   |||
 158 LysAlaGlySerValGlyMetGlyTyrProProGlnProIleSerGlyPh  174

2020 .CCCCATGGGCTAACAGCAGACGCGGGAGCTGGGA...CCCAGTGAGGCAG 1975
       ||| ||||||| |||  ::|||:::|||
 174 eProGlyProGlyProSerGlyProGlyProProGlyProGlyHisAlaG  191

1974 GCCCCTCCACCCAAATGCTGCAGAGTTTTCACGCTGAGTATTTGGCCAA 1925
      ||||||||  ||||
 191 lyProPro...................GlySer..............     195
```

```
835 ::|||::::::                   ::||||||
                                   ProGlnGlySerProGlyLe 845

567 CCCAAGGACAGTG...........CCCAGATGAAGGGCCG 539
              |||       ||| |
845 uProGlyInValGlyGluThrGlyLysLysGlyProProGlyArgAspG 862

538 GCGGGCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCG 489
                                ::|||||
862 ly..........ThrSerGlyLys.............Asp 867

488 GGACAGAGACCAGGCCCAGCACTGACCAATGCCCAGCCACCATGGTCATG 439
    ||:::|||  |||||||:::::::  |||:::
868 GlyGluArgGlyGlyProGlyValProGlyLeuProGly.......... 880

438 AACTTCTCCTCTACCCCCACTTCCAGCAGCAGCAGAGGCGGCACATAGGTGAT 389
    |||  ||  ::|||
881 ...LeuProGlyProValGlyProLysGlyGluProGly............ 892

388 GCCTCGCGGCCAAACACACCTCCAGCCCAAAGGTTAGCAGGTTGACCAGCA 339
    ||| :::|||  ||||||:::||  |||:::
893 ........ProValGlyAlaProGlyGlnValMetValGlyProProGly 906

338 AGAGCTGGCGTTTCCGGTGCC..........GCAGCAGGC.......GG 307
    ||||||| :::||||
907 AlaLysGlyGluGlyLysGlyAlaProGlyAspLeuAlaGlyAspLeuLeuGl 923

306 CTCACCCAGCCCTCTGGACCAGTAGTGGG.... 278
    :::||| |||:::  |||
923 yGluProGlyAlaLysGly.AspArgGlyLeuProGlyProArgGlyGlu 939

277 ..........CCAGGCCGGTAGGGCTAGGGG 255
    ||||||| :::||||
940 LysGlyGluAlaGlyArgGlyAlaGlyGlyLyGluProGlyAspProGl 956

254 CCGTTCAGGCCACTCCAGAACTGCTTCGTCTCGGC............. 221
    |::::::|||:::|||  ||
956 yGlnLysGlyAlaProGly......ArgLysGlyLeuLysGlyGluProG 971

220 ......TCTGCTCCAGAAGCTGCGGCGCTCTCCTTCGCTTTGCTGCCGC 182
    |||:::||  ::  :::|||||||
971 lyIleGluGlyValGlnGlyValProGlyProThrGlyProProGlyMetLys 987

182 ...................................... 182

988 GlyAspValGlySerProGlyAlaGlyProGlyValValGlyPheProGlyGl 1004

181 .CAACTGCCTAGG..........AATCAGCCAGGGCCCCATTTCTGCCAGC 142
    |||||| :::|||||||||| |||:::::
1004 nThrGlyProArgGlyThrGlyGlnProGly.ProValGlyGluArg 1020

141 CCTTTGGTG...........CCGGTCCAGCT 122
    ::  |||
1021 GlyLeuAlaGlyProProGlyArgGluGlyAlaProGlyPro........ 1034

121 TCTCAGCCATGCTCAACACCTGCTGCTGGGG......CACCTCAGTG 78
    ||| :::|||::::|||
1035 .LeuGlyProProGlyProGlySerValGlyAlaProGlyAlaSerG 1051

77 GGGACACGTCTCATCACTCAGATCCTGGC.............CGA 46
    |:: |||||||||:::||
1051 lyLeuLysGlyAspLysGlyAspProGlyThrGlyLeuProGlyProArg 1067

45 GGCGGCGGGCCTGTCACCCCGGA 25
    |||||| :::|||
1068 GlyGluArgGlyGlyProGly 1074
```

seq_name: pir2:I50694

seq_documentation_block:
collagen alpha 1(III) chain - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Jul-1998

```
1202 CTCTGGGCACGCCCTGTACAGCCCCTGCCCACGAAATCCGTGTAAAAC 1153
     :::|||              ::|||                    :|||
462  GluGluGlyLysArgGlyAlaAsnGlyAlaAsnGlyGluProGly.   473

1152 AGCGTGAAGGTCATGAGTGCCATCCAGCTGCCACAGCTCAGCCACGAAGAG 1103
     :::|||   |||||||::|||           |||       :::
474  .....GlnAsnGlyValProGlyThr.........ProGlyGluA       484

1102 CCGGGCCA.................GGTGCGGGGCATGCGGCCAGC 1074
     ||::::                          |||  :::  |||
484  rgGlySerProGlyPheArgGlyLeuProGlySerAsnGlyLeuProGly 500

1073 ACAGCTGGTGCGCCGGGAGCAGGGCGCCCAGGTTCCGGAAAGCCAAG 1024
     |||   |||||::|||   ||||| |||    |||||::::
501  GluLysGlyProAlaGlyGluArgGlySerProGlyProGlyProGlyProSe 517

1023 CGGGCCCGGATGGACAGCAGTGGGGCGCAAGGAGGGGGCCGACAGCCC 974
     ||||||:::                                :::
517  rGlyProAlaGlyAsp.............ArgGlyGlnAspGly.       527

973  TTCTGGCTCGGTGGGCCCCAGCCGCTGCCT..............       942
     ||||||:::||||||
528  ...............GlyProGlyLeuProGlyMetGlyArgGlyLeuPro 538

941  ..............................CCTCAGCCACCAGCA.      927
                                   ||||||||||
539  GlyIleProGlySerProGlySerAspGlyLysProGlyProProGlyAs  555

926  .GTGTGGCTGCTACGCAGGTGAGGAAGATGAGCGGTGAGCAGGCCAAAGAG 878
     |||   |||   |||||   |||   |||   |||   :::
555  nGlnGlyGluProGlyLysArgGlySerGlyProProGlyProAlaGlyProArgG 572

877  GCACTCCT.....................CCTGGGTGC           861
     ||  |||                      |||
572  lyGlnProGlyValMetGlyPheProGlyProLysGlyGluGluGlyAla  588

860  CCAGGT...........AGGGGCCCAGGCCACTGGTGTCCCAGTCAATG  823
     ||||||           |||||||||||
589  ProGlyLysAsnGlyLysArgGlyProGlyArgGlyProGlyThrProGl  605

822  GCAGGCAGGA...............GGTAGCCCAGGCAGCCCCCAAGACT  788
     ||||||                        ||||||||
605  yProAlaGlyLysAsnGlyAspArgGlyAspValGlyLeuProGlySerGlyProA 622

787  GATCATGAAGGCATAGACAGATAGGCCTGCCGACAGT...........  750
     :::|||||||::|||   |||||||||
622  laGlyProAlaGlyAspArgGlyLeuProGlyProSerProGlyLy      638

749  ..............GGTCCGGGTCCCGGAAAGGAGGTCAGAGAGCA..   720
                   ||||||   |||   :::|||
639  LeuGlnGlyLeuProGlyGlyProProAla...GlyGluAsnGlyLy     654

719  ...............................G                  719
654  sProGlyGluProGlyProLysGlyAspIleGlyGlyProGlyPheProG 671

718  GGCCTCCAGTGGAGTGAAGCACACCTGGCACAGAGTCCAGCAGCCCCA  669
     ||||||     ::|||   |||||||::|||
671  lyProLysGlyGluAsnGlyIleProGlyArgGlyLeuProGlyPro     687

668  .................CGCCCAGGATGAGCAGTGCCAGCT..        645
                      |||||||||
688  ProGlyProThrGlyAlaArgGlyProGly.........ProAlaGl    701

644  ...............................CCAGGGGCCTGGATCCGG  626
                                    |||||||||
701  yserGluGlyAlaLysGlyProGlyProGlyAla......ProG       716

625  GCACAGCAGCCCTGCTAGCCAGCCGCCCCTTGGGATGAGAAAGAGGCTCA 576
     :::||||||::|||     :::|||         |||||||||||
716  lyGlyThrGlyLeuProGlyLeuGlnGlyMetProGlyLeuArgGlyAla  732

575  GCAGGATGCCCAAGGACAGTGCCCAGGATGAAGGGCCCGGC.....GGCGG 532
     :::||||||   |||||||||                |||||
575  ysGlyGluArgGlyGlyAsnGlyThrProGlyLeuGlnGlyLeuArgGlyIla 461
```

```
1924 GTGCCTCTTGTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAGAC 1876
     ||||||                                      |||
196  ...............................AsnGlyTyrGlnGl       200

1875 ....CCAGGCCTGCGCAGACACCATAGGCAGTGACAGACTGGCTGAG   1829
     |||||        |||   ::
200  yProProGlyGluProGlyGlnProGlySerGlyProProGlyProAlaProA 217

1828 CTGGACAATGAGCCCATAAACAGGGATGGGCCCACTGGACACAGCAGGA 1779
     |||||::::  :::|||        |||||||||||
217  laGlyMetIleGlyPro........AlaGlyProProGlyLysGlyAspGly 230

1778 AGGCACTATCCAGGATGGCGAGGTCCAGGCAGATGCCCCGGCCGGAAC. 1730
     :::       ||||||  ||||| ||||||
231  Glu...............ProGlyArg..ProGly.ArgAsnG        239

1729 ...............................CACCCTGCCTC         1718
                                    ||||||:::
239  lyAspArgGlyIleProGlyLeuProGlyHisLysGlyHisProGlyMet 255

1717 GGTGGGCTCACC........................CACCACCA      1698
     ||||||:::
256  ProGlyMetProGlyMetLysGlyAlaAlaArgGlyPheAspGlyLysAspGl 272

1697 CACGTACGGAGACAATCACAGGCAGAGGCCCCGCAG..............  1663
     :::   |||||||   |||  ::|||:::
272  yAlaLysGlyAsp.SerGlyAlaAlaProGlyLysGlyGluAlaGlyGln  288

1662 ...AGCGCGGTGGAGGTGGGAGGCAGGCCACTGCCTCCAGCACCCACGTG  1616
     :::||||||||::::|||   |||   ::|||:::
289  ProGlyAlaAsnGlyGlyProGlyGlnProGlyProGlyProThrGl     305

1615 TCCATTAGGGAAG...GGAGCTCCAGGCTTAGGGCCT.......GGCA   1578
     |||||::::       |||  ||||| |||
305  yGluArgGlyArgProGlyAsnProGly...GlyProGlyAlaHisGlyL  321

1577 GGAAGCTGGTCATCAGGCTGTCCTCACTGCCTAGCACTCCAGTGTCCCT  1528
     ::  :::|||||||  :::|||:::||||||
321  ysAspGlyAlaProGlyThrAlaGlyProLeuGlyProProGlyProPro  337

1527 CGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGTAGAGGGGAG....   1482
     ||::::   |||||||   |||   |||  |||||
338  .GlyThr...AlaGlyPheProGlyAlaSerGlyArgGlyGluAlaG     353

1481 ......CCAGGTGTGTAGGGCGGAGAATCGCAGGCGTCGAGAGGTGA    1443
           ||::::   |||||||  ::|||   ||| |||||
353  lyProProGlyProAlaGlyAlaSerGlyAlaAsnProGlyLuArg..GlyGl 369

1442 ACCCGGTGAGGGCGGCTGAACGTCACCACGGCCACACTGTGGGACAGG   1393
     ||||||   :::|||   |||||| |||
369  uProGlyProGlnGly..GlnAlaGlyProProGlyProProGlyProPr  385

1392 CATGTGGCACCGGCACCGGGAAAGCTGCCACACTGCCCAAATAGAC     1343
     :::||||||  ||||| :::
385  oGlyArgAlaGlySerProGlyLysGlyLysGlyLysGlyGluMetGlyGlySerGlyI 402

1342 TGCTCCGAGTGCCGAATCGCTGCACCAGCCGCCCATGACCA.........  1302
     :::|||||| |||   ||::::|||
402  le..............ProGlyGlyProGlyProGlyProProGlyGly   411

1301 .............GAGAGAAGAACCAGGAGGAGATGGCG            1279
                  ::  :::|||||
412  ArgGlyLeuProGlyProProGlyThrSerGlyAsnProGlyAlaLysGl  428

1278 CACTGCCAGGAACAGCC...................CCAGGCTGCC    1253
     |||  |||||||         |||||::::
428  yThrGlyGluProGlyLysAsnGlyAlaAlaLysGlyAspProGlyProL  445

1252 CATCCGAACGCCTTCATCATAGTGTCTCCGGCGTCGTCCGGCCTCAG    1203
     :::||||||  :::   |||   |||   |||
445  ysGlyGluArgGlyGlyAsnGlyThrProGlyLeuThrProGlyLysGlyGly 461
```

```
                                 :::|||   |||              |||            |||
  733 SerGlySerProGly......ProLysGlyAspLysGlyGluProGlyGl   747

                                 |    ||| |||::          :::::::|
  531 CCATAGCGTCCAC............GCCAGTGGTCACTGGCTGAGCCTAG   494

                                 ||  |||::|    :::::|
  747 yLysGlyAlaAspGlyLeuProGlyAlaArgGlyGluArgGlyAsnValG   764

                                 ||||||||||||
  493 GAGCGGGACACAGACCAGCCAGCACTGGACCACCAATGCCCAGCACCATGG  444

                                 ||         :::::
  764 lyProIleGlyProGlyProProAla......GlyProProGlyAspLys   778

                                 ||                  ||
  443 TCATGAACTTCTCCTCTA...CCCCCACTTCCAGCAGCAGAGGGGCCACA   397

                                 :::::|||  |||||||:::::: :::
  779 GlyGluThrGlyProAlaGlyAlaProGlyAlaProGlySerArgGly..  794

                                 |||||||:::|||
  396 TAGGTGATGCGCGGGCCAAACACCACCTCCAGGCCAAAGGTTAGCAGGTT  347

                                 |||||||:::|||
  795 ...............GlyProGlyGluArg...........          799

                                 ::::::   |||  ||||||||  |||::
  346 GACCAGCAAGAGCTGGGCTTTCCGGTGCCCAGCAGGCGGCTCACCCACA   297

                                 :::::   ||||||   ||||:::::
  800 ....GlyGluGlnGlyLeuProGlyGlyProAla.....GlyPheProGly  812

                                 |||          |||||||   |||
  296 GCCTCTGGACCATAGTGGG...CCAGGCGGGTAGGGCTCAGGGGCCGTT   250

                                 |||  :::|||||:|||   |||||||||  |
  813 AlaProGly.GlnAsnGlyGluProGlyLyGlyLysGlyGluArgGlyProP 829

                                 ||  ||||||:::||||::||
  249 CAGGCACTCCAGAAACTGCTTCGTCGGGTCTGCTCCAGAAGCTGCGGCC   200

                                 |||       |||||:::|||||||
  829 roGly.........LeuArgGlyGluAlaGlyProProGlyAlaAla     841

                                 :::|||      ||||||
  199 TCTCCCTCCTTGCTGCGCCCAACTGCCTAGGAATCAGCCAGGCGCGCCATT 150

                                 :::            |||||||  |||
  842 Gly.......................ProGlnGlyIleGlyProGlyAla.....  849

                                 ||||||
  149 CTTGCCAGCGCCCTTGGGTGCCGGCCAGCTTCCAGCCCATGCTCAACACCT  100

                                 ||||||
  850 ...........ProGlyProPro............................  853

                                 |||||||
   99 GCTGCTGTGGGCCACCTCAGTGGGGACACGTCTCATCACTCAGATCCT..   52

                                 :::|||||| ||||||| :::::::|:::|||
  854 ...........GlyProGln.GlyValLysGlyGluArgGlySerProGl   866

                                 .:::||||||
   51 .GGCCGAGGCGGCGGCGGCCGTGTCACCCGGAGCC     22

                                 ||||||||
  866 yGlyProGlyAlaAlaAlaGlyPheProGlyAla     876
```

```
seq_name: pirl:CGHU1V

seq_documentation_block:
collagen alpha 1(V) chain precursor - human
N;Alternate names: procollagen alpha 1(V) chain
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence_revision 03-Oct-1995 #text_change 26-Feb-1999
C;Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665
R;Greenspan, D.S.; Cheng, W.; Hoffman, G.G.
J. Biol. Chem. 266, 24727-24733, 1991
A;Title: The pro-alpha1(V) collagen chain.  Complete primary structure, distribution of e
A;Reference number: S18802; MUID:92105142
A;Accession: S18802
A;Molecule type: mRNA
A;Residues: 1-1838 <GRE>
A;Cross-references: GB:M76729; NID:g189519; PID:g189520
R;Takahara, K.; Sato, Y.; Okazawa, K.; Okamoto, N.; Noda, A.; Yaoi, Y.; Kato, I.
J. Biol. Chem. 266, 13124-13129, 1991
A;Title: Complete primary structure of human collagen alpha-1(V) chain.
A;Reference number: S16024; MUID:91302336
A;Accession: S16024
A;Molecule type: mRNA
A;Residues: 1-81,'QL',84-389,'A',391-676,'K',678-1294,'PS',1297,'RS',1300-1553,'R',1555-
A;Cross-references: GB:D90279; NID:g219509; PID:d1015029; PID:g219510
A;Note: parts of this sequence were determined by protein sequencing
R;Yaoi, Y.; Hashimoto, K.; Takahara, K.; Kato, I.
```

```
Exp. Cell Res. 194, 180-185, 1991
A;Title: Insulin binds to type V collagen with retention of mitogenic activity.
A;Reference number: A61142
A;Accession: A61142
A;Molecule type: protein
A;Residues: 823-824,'X',826-842 <YAO>
A;Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyano
R;Yaoi, Y.; Hashimoto, K.; Koltabashi, H.; Takahara, K.; Ito, M.; Kato, I.
Biochim. Biophys. Acta 1035, 139-145, 1990
A;Title: Primary structure of the heparin-binding site of type V collagen.
A;Reference number: S11303; MUID:90366601
A;Accession: S11303
A;Molecule type: protein
A;Residues: 823-824,'X',826-848,'I',850-851,'P',853,'PR',856-893,'D',895-932,'X',934-
A;Note: the residues designated 'X' are probably glycosylated hydroxylysine; this seq
R;Seyer, J.M.; Kang, A.H.
Arch. Biochem. Biophys. 271, 120-129, 1989
A;Title: Covalent structure of collagen: amino acid sequence of three cyanogen bromid
A;Reference number: S03978; MUID:89927189
A;Accession: S03978
A;Molecule type: protein
A;Residues: 621-640,'G',642-649,'L',651-662,'E',664-667,'Q',669-676,'Q',678-683,'P',6
<SEY>
A;Note: there are a number of inconsistencies between the sequences in figures 6 and
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champliaud, M.F.; Boutillon, M.M.;
Eur. J. Biochem. 221, 987-995, 1994
A;Title: Diversity in the processing events at the N-terminus of type-V collagen.
A;Reference number: S43642; MUID:94237164
A;Accession: S43642
A;Molecule type: protein
A;Residues: 565-576;755-758,'X',760-763,'X',765-772;1012-1029;1219-1232;1465-1474,'X'
R;Fessler, L.I.; Brosh, S.; Chapin, S.; Fessler, J.H.
J. Biol. Chem. 261, 5034-5040, 1986
A;Title: Tyrosine sulfation in precursors of collagen V.
A;Reference number: A56977
A;Contents: annotation; identification of tyrosine sulfate in the amino-terminal prop
R;Lee, S.; Greenspan, D.S.
Biochem. J. 310, 15-22, 1995
A;Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).
A;Reference number: S58665
A;Accession: S58665
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-36 <LEE>
C;Comment: Prolines and lysines are at the third position of the tripeptide repeating uni
C;Comment: the heterotrimers are probably processed to the long form.
C;Comment: A long form of the mature protein containing part of the amino-terminal pr
are 5-hydroxylated and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL5A1
A;Cross-references: GDB:131457; OMIM:120215
A;Map position: 9q34.2-9q34.3
C;Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer o
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among
ength, is formed with desmosine cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with
A;Note: may play a role in controlling the lateral growth of collagen I fibrils
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homolo
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr
F;1-37/Domain: signal sequence #status predicted <SIG>
F;36-261/Domain: PARP-like #status predicted <PARP>
F;38-541/Domain: amino-terminal propeptide #status predicted <PRO>
F;38-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>
F;542-558/Region: amino-terminal nonhelical telopeptide
F;559-1572/Region: helical
F;645-647/Region: cell attachment (R-G-D) motif
F;663-665/Region: cell attachment (R-G-D) motif
F;897-929/Region: heparin binding
F;1573-1605/Region: carboxyl-terminal nonhelical telopeptide
F;1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic
```

F;62-244,183-237/Disulfide bonds: #status predicted
F;159,176,385,1672,1741/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;234,236,240,262,263,274,275,277,279,280,338,340,346,347,352,357,416,417,420,421/Bi
F;535/Modified site: allysine (Lys) #status predicted
F;541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F;542/Modified site: Pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;570,576,621,639,648,654,657,675,678,690,693,696,705,717,720,726,732,741,750,753,756,76
site: 4-hydroxyproline (Pro) #status experimental
F;627,642,687,708,744,774,795,804,807,810,819,825,846,864,882,897/Modified site: 5-hydro
F;627,642,687,774,795,800,810,819,825,846,864,882,897,1482/Binding site: carbohydrat
F;708,744/Binding site: carbohydrate (Lys) (covalent) #status experimental
F;1482/Modified site: 5-hydroxylysine (Lys) #status predicted
F;1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted
F;1639,1645,1662,1671/Disulfide bonds: interchain #status predicted
F;1680-1835,1746-1789/Disulfide bonds: #status predicted

alignment_scores:
        Quality: 342.00   Length: 1169
        Ratio: 0.742   Gaps: 78
Percent Similarity: 39.435   Percent Identity: 26.604

alignment_block:
US-09-030-606-110/rev x CGHU1v   ..

Align seg 1/1 to: CGHU1v from: 1 to: 1838

3217 AAACGGCACTTAAACCCCCCCTGAGAGATAAGACCTCCCTTAGCTCAG.. 3170
     ||| ::||||||:|| ||||| |||||
708 LysGlyAsnValGlyProGlnGlyGluProGlyProGlnGlyGlnGlnGlnI 724

3169 .........GCAGGGGGTGCTCCTGAGTTTCTGTGTGAGATTCCCCAAG 3130
             ||| || ||| |||||||
724 yAsnProGlyAlaGlnGlyLeuProGly............ProGlnG 736

3129 CACAGATATACTCTGGGGCTGAGATGACAAAGGCTTGGGAAACCGCAC 3080
     :: ||| ||||||||||||
736 lyAlaIleGlyProProGlyGluLuLysGlyPro...LeuGlyLysProGly 751

3079 TTTGT........GCTTCTGTCCTCGCAGTAGCTCCAAACAGGGT 3043
     ::| ||||:::||||| |||:::||:
752 LeuProGlyMetProGlyAlaAspGlyProProGlyHisProGlyLysGl 768

3042 TGTGGAGCTGGTGGGGAAAGTTGGGGTGGTAGGGGGAAAGTTGGGGTAGGGG 2993
     ||| ||||||::: || ||| || |||||||| ||| :::||
768 u..GlyProProGlyGluLuLysGlyGlyGlnGlyProProGlyProGlnGl 784

2992 AAATTTTGGGCAGTGCCTTCATCAGCCCAGTCCTAGAGAGAGAGGG. 2944
     | ||||||:||||||| ||| :||||||||:::::
784 yProIleGly........TyrProGlyProArgGlyValLysGlyA 797

2943 ................................GAGTGGAAGTGG 2932
                                      :::   :: |||
797 laAspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAsp 813

2931 GGGGAACCAGGCTGGCCAAGAGAAGAGGGTGGTTAGGGAAGCCGTTGA 2882
     ||| |||||||::: |||::   ::: ||| ||
797 laAspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAsp 813

2881 GACCTGAAGCCTCACCCTCTACCTTCCTTCAACACCCTAACCTTGGGTAA 2832
     ||| ||||:: ||||||||::::|||| |||:|||| ::|||||
830 uIleGlyProPro............. 834

2831 CAGCATTTGGAATTATCATTTGGATGAGTAGTAGAAATTCCAAGGTCCTGGG 2782
     |||||| ::||| |||||
835 ....GlyPro.... 836

2781 TTAGGGCATTTTGGGGGCCAGACCCCCAGGAGAAGAGATTCTGGCCAATGA 2732
     ||| ||| :::|||||||::||||:|| |||::||
837 ....ArgGlyGluAspGlyProGluGlyProLysGlyArgGl 849

2731 TCAGCCCAATGA........................CCAGCTATCTCAGGGG 2704
     |||||| |||:: ::|||

Page 32

us-09-030-606-110.rpr

Wed Sep 29 14:26:56 1999

1059 CGGGGAAGCAGGGCGGGCCCAGGTTCCGGGAAAGCGGGAAGCGGCCCGGCCATGG 1010
|||||| ||| ::::||||| ||| ||||||
1290 aGlyGluProGlyLeuProGlyGluGlyLeuProGlyProGlyProLysGlyG 1307

1009 ACAGCAGTGGGGCGACAAGGAGGGGCCGACAGCCCTTCGCTGGCTCGG 960
::: ::: ||||||:::::
1307 luArgGlyGluLysGlyLeuSerGlyProSerGlyAlaAla....... 1320

959 TGGGGCCAGGCTGCCTCCTCAGCCACCAGCAGTGTGGCTGCTACGCAG 910
|||||| ||| |||||
1321 ..GlyProProGlyProLysProGlyProPro............. 1329

909 GTGAGGAAGATGAGGTGAGCAGGCCAAAGAGGCACTCCT.......... 870
||| ||||:::|||:::|||
1330 ....GlyAspAspGlyProLysGlySerProGlyProValG 1342

869 ....CCTGGGTGCCCAGGTAGGGGGCCAGGGCACTGGTGTCCCAGTCAA 825
|||||| |||:::|||
1342 lyPheProGlyAspProGly...ProProGlyGluProGlyProAla... 1356

824 TGGCAGGCAGGAGGTAGCCCAGGCAGCCCCAAGACTGATCATGAAGGCA 775
||||: ||| ||| :::
1357 ....GlyGlnAspGlyProProGlyAsp........LysGl 1366

774 TAGACAGAGTAGGCCTGGCGACAGTGGT...CCGGGTCCCGGAAGAGGTC 728
:::|||||||||:::||||||| |||||||
1366 yAspAspGlyGluProGlyGlnThrGlySerProGlyPro........ 1379

727 AGAGACAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCA 678
::|||| ||| ||||||| |||:::|||
1380 .....ThrGlyGluProGlySerGlyProProGlyLysArgGlyPro 1394

677 GCAGCCCCACGCCCAGGATGAGCAGTGGCCAGCTCCAGGGGCCTGGGATCC 628
||| ||| :::
1395 ....Pro.Gly.......... 1396

627 GGGCCACAGCCAGGCCCTCCTAGCCACCGGCCCCTTGGGATGAGAAAGAGGCT 578
|||||||||| ::: :::||||||
1397 ....ProAlaGlyPro...GluGlYArgGlnGly 1405

577 CAGCAGGATGCCCAAGGACAGTGCCCAGTGAAGGGGCCGGCGGCCCAT 528
:::::: |||::||||||| ||||||||
1406 GluLysGlyAlaLysGlyGluAlaGlyLeuGluGlyPro.......... 1418

527 AGCGTCCACGCCAGTGGTCACTGCTG....... AGCCTAGGAGCGGGAC 485
||| |||||||| ||||||
1419 .....ProGlyLysThrGlyProIleGlyProGlnGlyAlaProG 1432

484 ACAGACCAGGCCCA...GCACTGACCAATGCCCAGCACCATGTCATGA 438
::|||||||| :::||| |||:::|||
1432 lyLysProGlyProAspGlyLeuArgGlyIleProGlyProValGly.. 1447

437 ACTTCTCCTCTACCCCACTTCAGCAGCAGGCGGCACATAGGTGATG 388
:::||| |||||
1448 GluGlnGlyLeuProGlySerProGly............... 1456

387 CCTGGCGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCAA 338
:::||| ||| ||| :::|||
1457 ......ProAspGlyProProGlyPro......MetGlyProProGlyL 1469

337 GAGCTGGCTTCCGTGCCCAGCAGGC...................... 309
||||||:::||
1469 euProGlyLeuLysGlyAspSerGlyProLysGlyGluLysGlyHisPro 1485

308 .....GGCTCACCCACCTGGACCATAGTGGG.. 278
||| ||| ||||
1486 GlyLeuIleGlyLeuIleGlyProProGlyGluGlnGly.GluLysGlyA 1502

277 ..........CCAGGCGGTAGGGCTCAGGGGCCGTTCAGGCACTCCA 239
||||||:::|||
1502 spArgGlyLeuProGlyGlnGlyProGlySerSerGlyProLysGlyGluGln 1518

238 GAACTGCTTGCTTCGTCTCGGCCTCGCTGCTCCAGAAGCTGCGGCCTCTCCTTG 189

1836 TGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATGGGGCCACTGGGA 1787
||||||||
1051 ....GlyProProGlyL 1055

1786 CAGCAGGAAGGCACTATCCAGGATGGCGAGGTCCAGGCAGATGCCCCGGC 1737
:::||||| ||||| |||||
1055 euArgGly.....PheProGlyAspPArgGly........LeuProGly 1066

1736 ....CCGGAACCACC 1726
||| |||||
1067 ProValGlyAlaLeuGlyLeuLysGlyAsnGluGlyProProGlyProPr 1083

1725 CTGGCCTCGGTGGGCTCACCCACCACCACAGTACGGAGACAATCACAGGC 1676
:::||||||||
1083 oGlyPro..AlaGlySerPro..........GlyGluArgGlyProAl 1095

1675 AGAGGCCCCCAGGAGCCGGGGTGGAGTTGGGAGGCAGGGCCACTGCCTCCAG 1626
||| ||| |||
1095 aGlyAlaAlaGlyProIleGlyIleProGly..ArgProGlyProGlnG 1111

1625 CACCCACGTGTCATTAGGG...AAGGGAGCTCCAGGC.......... 1591
::||| ||| |||||||||||
1111 lyProProGlyProAlaGlyGluLysGlyValAlaProGlyValGlyLysGlyPro 1127

1590 TTAGGGCCTGGCAGGAAGTGGTCATCAGCGTGTCCTCACTGCTAGCACC 1541
|||||||| :::|||
1128 GlnGlyPro.AlaGly......ArgAspGly.............L 1136

1540 TCCAGTGTCCCCTCGGTATTTGGGCAGGACAACCTGCTTCTCCGGTGGT 1491
||||| |||:::|||:::
1136 euGlnGlyProValGlyLeu......ProGlyProAlaGlyProValGly 1150

1490 AGAGGGAGGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAAC 1441
||| :::|||
1151 ....ProProGlyGluAspGlyAsp.LysGlyIleGluGlyGluP 1164

1440 CCGGTGAGGGGCGGCTGAAGCTGTCACCACGGCCAC.......... 1406
::: ||||: |||:::
1164 roGlyGlnLysGlySerLysGlyAspLysGlyLysGlyGlnGlyProProGly 1180

1405 ...............ACTGTGGACAGGCATGTGGCCACGGCAGCCA 1374
::|||||| ::|||:::
1181 ProThrGlyProIleGlyGlnProGlyIleProGlyProSerGlyAlaAs 1197

1373 CAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGC 1324
||| ||| |||||
1197 pGlyGluProGlyProArgGlyGln.................... 1205

1323 TGCACCAGCCGCCGGTCCATGACCAGAGAAGACCAGGGAGATGGC...... 1280
||| :::||| |||:::|||
1206 .....GlnGlyLeuPheGlyGlnLysGlyAspGluGlyProArgGlyPhe 1220

1279 ..........GCACTGCAGGAACAGCCCCAGGGTGCCCCA 1251
::||||||| |||||
1221 ProGlyProProGlyProValGlyLeuGln..GlyLeuProGlyProPro 1236

1250 TCCGAACGCCTTCATCATAGTGTCTCCGGCCTCGGTGCCCGGCTCAGCT 1201
:: ||| |||
1237 GlyLysLysGlyGluThrGlyAspValGlyGlnMetGlyPro........ 1250

1200 CTGGGCACGCCCTGTACAGCCCCTGCGCCACGAAATCCGTGTAAAACAG 1151
|||||| ||| ||||||:::|||
1251 ....ProGly...ProProGlyProArgGlyPro.......... 1259

1150 CGTGAAGGTCATGAGTGCCATCAGCTGCACAGCTCAGACCACGAAGAGCC 1101
||| :::||||:::||:::
1260 ....SerGlyAlaProGlyAlaAspProGlyGlnGlyProProGly 1273

1100 GGGCCAGGG...TGCCGGCATGGCGGCAGCAGT.....GGTGCAGC 1060
||| ||| |||
1273 ProGly...ProProGlyProArgGlyPro.......... 1259...

1100 GGGCCAGGG...TGCCGGCATGGCGGCAGCAGT.....GGTGCAGC 1060
||| ::: ||| ||
1274 GlyIleGlyAsnProGlyAlaValGlyLysGlyGluProGlyGluAl 1290

```
1519 GlyIle..........          ::::::||||::::::   ::|||||||
                               ThrGlyProSerGlyProIleGlyProProGl 1531

 188 CTGCCCCAACTGCCTAGGAATCAGCCAGGCGCCATTCTGCCAGCCCT 139
                     ||||||    :::  ||||||    |||
1531 yProProGlyLeuPro..GlyProProGly............ProL 1542

 138 TTGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGGG  89
     ||||||          :::|||  |||
1542 ysGlyAlaLysGlySerGlyProThr..................... 1551

  88 GCACCTCAGTGGGGACACGTCTCATCACTCAGATCCTGGC  49
     ::||||::  ||  :::::|||   :::    ||||||
1552 GlyProLys.GlyGluAlaGlyHisProGlyProProGly 1564
```

seq_name: pir2:JQ2389

seq_documentation_block:
sucrose transport protein - potato
C;Species: Solanum tuberosum (potato)
C;Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 09-Sep-1997
C;Accession: JQ2389; S40310
R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.
Plant Cell 5, 1591-1598, 1993
A;Title: Potato sucrose transporter expression in minor veins indicates a role in phloem
A;Reference number: JQ2389; MUID:94146554
A;Accession: JQ2389
A;Molecule type: mRNA
A;Residues: 1-516 <RIE>
A;Cross-references: EMBL:X69165; NID:g439293; PID:g439294
A;Experimental source: cv. Desiree
C;Comment: The gene encoding for this protein is highly expressed in mature leaves.
C;Keywords: glycoprotein; transmembrane protein
F;31-53/Domain: transmembrane #status predicted <TM1>
F;67-86/Domain: transmembrane #status predicted <TM2>
F;103-122/Domain: transmembrane #status predicted <TM3>
F;141-160/Domain: transmembrane #status predicted <TM4>
F;180-200/Domain: transmembrane #status predicted <TM5>
F;226-248/Domain: transmembrane #status predicted <TM6>
F;285-304/Domain: transmembrane #status predicted <TM7>
F;331-349/Domain: transmembrane #status predicted <TM8>
F;366-385/Domain: transmembrane #status predicted <TM9>
F;409-427/Domain: transmembrane #status predicted <TM10>
F;429-448/Domain: transmembrane #status predicted <TM11>
F;3,92/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
          Quality: 333.00   Length: 539
          Ratio: 1.198   Gaps: 21
Percent Similarity: 51.577   Percent Identity: 24.675

alignment_block:
US-09-030-606-110 x JQ2389

Align seg 1/1 to: JQ2389 from: 1 to: 516

```
 278 CCCACTATGGTCCAGAGGCTGTGGGTGGGTGAGCCGCTGCTGCGGCACCGGAA 327
         |||                :::::|||||||
  24 ProLeuAlaProSerLysLeuTrp...................... 31

 328 AGCCCAGCTCTTGCTGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTT 377
     :::::::||||| |||         ||||:::
  32 ...LysIleIleValAlaSerIleAlaAlaGlyValGlnPheLysAlaGlyT 47

 478 GGTGCTGTGTCCGGTCCTCTAGGCTCAGCCAGTGACCACTGGCTGGACGCT 527
     ::::::|||::::::||||       |||||||||   :::|||
  80 tIleValGlnProValValGlyTyrTyrSerAspAsnCysSerSerArgP 97

 528 ATGGCGCCGCCGGCGCCCTTCATCTGGCACGTCCTTGGGCATCCTGCTG 577
     :||||||||||||||||||| |||          ||
  97 heGlyArgArgProPheIleAlaAlaGlyAlaAlaLeuValMetIle 113

 578 AGCCTTTTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTG....... 619
     ::||||||| |||           |||||
 114 AlaValPheLeuIle.......GlyPheAlaAlaAspLeuGlyHisAl 127

 620 ..........CTGTGCCCGGATCCAGGCCCCTGGAGCTGGCACTGC 656
                   |||         :::|||
 127 aSerGlyAspThrLeuGlyLysGlyPheLysGlyProArgAlaIleAlaValP 144

 657 TCATCCTGGCGTGGGGCTGCTGGACTTCTGTGGCCAGGTGTGCTTCACT 706
     :::||||||      ::|||||
 144 heValValGlyPheTrpIleLeuAspValAlaAsnAsnMetLeuGlnGly 160

 707 CCACTGGAGGCCCGCTCTCTGACCTTCCGGGACCGGAC...CACTG 753
     |||          |||         :::||||||::|||||
 161 ProCysArgAlaLeuLeuAlaAspLeuSerGlyGlyLysSerGlyArgMe 177

 754 TCGCCAGGCCTACTCTGCTATGCCTCATGATCAGTCTTGGGGCTGCC 803
     |||  |||     :::|||::::::|||:::  |::::|||:::  |
 177 tArgThrAlaAsnAlaPhePheSerPhePheMetAlaValGlyAsnIleL 194

 804 TGGGCTACCCTCCTGCCTGCCATTGACTGG................ 832
     ||||||||
 194 euGlyTyrAlaAlaGlySerTyrSerHisLeuPheLysValPheProPhe 210

 833 ...GACACCAGTGCCCTGGCCCCTACCTGGCACCCAGGAGGAGTGCCT 879
     ::|||||||      |||::||
 211 SerLysThrLysAlaCysAspMetTyrCysAlaAsnLeuLysSerCysPh 227

 880 CTTT...GGCCTGCTCACCCTCATCTTCCACCTCGGTAGCAGCCACAC 926
     ::|||       |||:::  ||||||   |||    :::||||||
 227 ePheIleAlaIlePheLeuSerLeuThrThrIleAlaLeuThrL 244

 927 TGCTGGTGGCTGAGGAG..GCAGCCGTGGGCCCACCGAGCCAGCAGAA 973
     |:::               ||| :::|||
 244 euValArgGluAsnGluLeuProGluLysAspGluGlnIleAspGlu 260

 974 GGGGCTGCGGCCCCCTTGTCGCCCCACTGCGTCCATGCGGGCCCG 1023
     ||:::||||||                         :::
 261 LysLeuLeuAlaGly...........................LysSerLy 269

1024 CTTGGCTTCCGGAACCTGGGCCGCCCGCTTCCCGCGCCTGCACCAGCTGT 1073
     :::  |||       ::|||        |||     |||||||||
 269 sValProPhe.....PheGlyGlyIlePheGlyAlaLeuLysGluLeu. 283

1074 GCTGCCGCATGGCCACCTCATGACCTTCACGCTGTTTTACGCGGATTTCGTTGGG 1123
     |||::::    ||||||       |||:::|||::::||||||||||
 284 ...........ProArgProMetTrpIleLeuLeuLeuValThrCysLeu 296

1124 AGCTGGATGGCACTCATGACCTTCGGATGGGGCAGCCTGGGCGTTCCTGG 1173
     ::|||||    |||:::||        ||||||::::::|||||||||||
 297 AsnTrpIleAlaTrpPhePhePheLeuTyrAspThrAspTrpMetAl 313

1174 CGAGGGGCCGCATGGCCGTGCCCAGAGCGTGCCCGGCCACCGAGGCCC 1223
     :::  :::  ||||          |||||||||||||
 313 aLysGluValPheGlyGly..........GlnValGlyAsp.......A 324

1224 GGAGAACTATGATGAAGGCGGTTCGGATGGGCAGCCTGGGCGTGTTCCTG 1273
     |||  ||||||   |||||||||       |::::::||||||||||
 324 laArgLeuTyrAspLeuGlyIleValArgAlaGlyValAlaMetGlyLeuLeuLeu 340

1274 CAGTGCGCCATCTCCCTGGTCTTCTCTCGTGGTCATGGACCGGCTGTGCA 1323
     |||       :::||:::||||||  :::||||||
 341 GlnSerValLeuLeuPheMetSerLeuGlyValGluPheLeuGlyLy 357

1324 GCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTG 1373
```

```
 378 TGGCCAGGCAGCATCACCTATGTGCCGCCTCTTGCCGCTGGAAGTGGGGGTA 427
         |||||::||         ||          |::::::|||||||:::|::
  47 rpAlaLeuGlnLeuLeuSerLeuThrProTyrValGlnLeuLeuGlyIle 63

 428 GAGGAGAAGTTCATGACCATGTGCTGGGCCATTGGTCCAGTGCTGGGCCT 477
         ||||||||||  :::||||:::  :::  |||  :::   |||
  64 ProHisLysPheAlaSerPheIleTrpLeuCysGlyProIleSerGlyMe 80
```

A;Molecule type: mRNA
A;Residues: 403-419,'GR',421-536,'P',538-846,'VM' <KIM>
A;Cross-references: EMBL:X14963; NID:g29984; PID:g930048
C;Comment: Prolines and lysines at the third position of the tripeptide repeating uni
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL16A1
A;Cross-references: GDB:134045; OMIM:120326
A;Map position: 1p34-1p34
C;Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XV
C;Function:
A;Description: structural component of extracellular fibrous polymer as a minor form
A;Note: may play a role in forming elastic connections at fibril surfaces
C;Superfamily: unassigned collagens
C;Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxylys
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-333/Product: collagen alpha 1(XVI) chain #status predicted <MAT>
F;22-333/Domain: amino-terminal nonhelical #status predicted <NC11>
F;334-1577/Region: interrupted helical
F;334-360/Domain: collagenous COL10 #status predicted <CO10>
F;375-505/Domain: collagenous COL9 #status predicted <COL9>
F;521-554/Domain: collagenous COL8 #status predicted <COL8>
F;539-541/Region: cell attachment (R-G-D) motif
F;572-630/Domain: collagenous COL7 #status predicted <COL7>
F;652-722/Domain: collagenous COL6 #status predicted <COL6>
F;738-875/Domain: collagenous COL5 #status predicted <COL5>
F;887-938/Domain: collagenous COL4 #status predicted <COL4>
F;973-987/Domain: collagenous COL3 #status predicted <COL3>
F;1005-1007/Region: cell attachment (R-G-D) motif
F;1011-1432/Domain: collagenous COL2 #status predicted <COL2>
F;1226-1228/Region: cell attachment (R-G-D) motif
F;1472-1577/Domain: collagenous COL1 #status predicted <COL1>
F;1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>
F;47,327/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
        Quality: 330.00   Length: 1262
        Ratio: 0.643   Gaps: 81
Percent Similarity: 40.650   Percent Identity: 25.040

alignment_block:
US-09-030-606-110/rev x S23810   ..

Align seg 1/1 to: S23810 from: 1 to: 1603

3077 TGTGCTTCTTGGTCCT...............................GCAGTAGCTCC 3052
     |||||:::||||||| ||:|||||||||
 372 CysAlaGluGlyProLysGlyGluLysGlyGlyuSerGlyAlaLeuGlyPr 388

3051 AAACAGGGTTGTGGAGCTGGTGGGGGAGAAGTTGGGGTAGGGGAAAGTTGG 3002
     |::: ||| :::||||::: ||| ::||||::   ::||||::
 388 oSerGlyLeuProGlySerThrGlyGluLysGlyGlnLysGlyGluLysG 405

3001 GGGTAGGGGAAATTTGGGCAGTGCCTTCATCAGCCAGTCCTAGAGAGA 2952
     || ||| ||| ||| |||||| ::||||: ::||||: |||:||
 405 lyAspGlyGlyIleLysGly.ValProGly.LysProGly...ArgAspA 420

2951 GTAGAGGGGAG...............................TGGAAGTGGGGGAACCA 2923
     ::|||| |||||||||||||||
 420 laProGlyGluIleCysValIleGlyThrGlyGlnLysGlyAspPro 436

2922 GGCTGGGCCAAGAGAGAGAGGGTGGTTAGGGAAGCCGTTGAGACCTGAAG 2873
     ||:::||:::||:::|||:::|| ::
 437 GlyPheValGlyProGluGlySerThrGlyLeuAlaGlyGluPro.........GI 449

2872 CCCCACCCTCTACCTTCCTTCAACACCCTAACCTTGGGTAACACAGCATTTG 2823
     ||| ||| |||
 449 yPro........ProGlyLeuProGlyPro........................ 456

2822 GAATTATCATTTGGGATGGATGAGATAGAATTTCCAAGGTCCTGGGTTAGGCCATT 2773
     |||||||:::|||::: ||||||||:::||:::|||:||
 457 ....ProGlyIleGlyLeu 461

---

357 :::::|| ||| ||| :::::|||
    sLysIleGly...GlyAlaLysArgLeuTrpGlyIleLeuAsnPheValL 373

1374 TGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTG...... 1417
     :: ||| ||||||:::   :::::||||::
 373 eu.......AlaIleCysLeuAlaMetThrIleLeuValThrLysMet 386

1418 .........ACAGCTTCAGCCGCCGCCCTCACCGGGTTCAC 1446
          :::::::||| ||| ||
 387 AlaGluLysSerArgGlnHisAspProAlaGlyThrLeuMetGlyProTh 403

1447 .   TTCTCAGCCCTGCAGATC. 1465
     |- |||:::|||||| |||
 403 rProGlyValLysIleGlyAlaLeuLeuLeuPheAlaAlaLeuGlyIleP 420

1466 .   CTGCCCTACACACTGGCCTCCCCTCTACCAC 1495
     :::||||:::|||||||||:::::
 420 roLeuAlaAlaThrPheSerIleProPheAlaLeuAlaSerIlePheSer 436

1496 CGGGAGAAG...............................CA 1506
     :::::
 437 SerAsnArgGlySerGlyGlnGlyLeuSerLeuGlyValLeuAsnLeuAl 453

1507 GGTGTTCCTGCCCAAATACCGAGGGGACACCTGGAGGTGCTAGCAGTGAGG 1556
     :: :::|||::: ||| :::|| :::|
 453 aIleValAlaProGlnMetLeuVallSerLeuValGlyGlyProTrpPAspA 470

1557 ACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTC 1606
     || ::: ::: |||||||||
 470 spLeuPheGlyGlyLysAsnLeuProGly.......................Phe 480

1607 CCTAATGGACACGTGGGTGCTGGAGGCAGTGGC...............CT 1641
     ||| :::|||:::||:::||||||||
 481 ValAlaGlyAlaValAlaAlaAlaAlaSerAlaValAlaLeuAlaLeuThrMe 497

1642 GCTCCCACCTCCACCCGCGGCCTCTGCGGCGGCCCTCGCGCGGCCCTGGTG 1691
     :|||||| ||:::::
 497 tLeuProSerProProAla..............................AspAlaLysP 507

1692 TACGTGTGGTGGTGGGT 1708
     ||:::::|||
 507 roAlaValAlaMetGly 512

seq_name: pir2:S23810

seq_documentation_block:
collagen alpha 1(XVI) chain precursor - human
N;Alternate names: procollagen alpha 1(XVI) chain
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 24-Sep-1998
C;Accession: S23810; PQ0612; S08012
R;Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.
A;Title: Cloning and chromosomal location of human alpha1(XVI) collagen.
A;Reference-number: S23810; MUID:92335339
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992
A;Accession: S23810
A;Molecule type: mRNA
A;Residues: 1-1603 <PAN>
A;Cross-references: EMBL:M92642; NID:g180757; PID:g180758
A;Experimental source: skin fibroblasts
R;Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakosh
J. Biochem. 112, 855-863, 1992
A;Title: Molecular cloning and partial characterization of a novel collagen chain, alpha
A;Reference number: PQ0612
A;Accession: PQ0612
A;Molecule type: mRNA
A;Residues: 'GGR',421-536,'P',538-1159,'P',1161-1162,'P',1164,'P',1166-1603 <YAM>
A;Experimental source: placenta
R;Kimura, S.
submitted to the EMBL Data Library, April 1989
A;Description: Partial nucleotide and amino acid sequence of a collagen-like protein fro
A;Reference number: S08012
A;Accession: S08012

us-09-030-606-110.rpr

```
2065 CCACATTACTTGGCAGCAACAGAGAAACTGGCGGCCCGGCCAGCCCC.    2017
     ||||| ::: :::|:::::||:::||::||||
 677 ArgGlyLeuLysGlyGlnLysGlyAsp.AlaGlyAsnProGlyAspProGly  693

2016 .....ATGGGCTAACAGGAGCGGGG..........AGTGGGACCCA...GTG  1981
          ||| |||||| |||    :::||| |||
 693 lyThrProGlyThrThrGlyArgProGlyLeuSerGlyGluProGlyVal   709

1980 AGGCAGGCCCTCACCCCAATGTGCTGAAGTTTCTACGCTGATATTT      1931
     ::: ||||:: |||  :::|||:::|||||:::
 710 Gln.GlyProAlaGlyProLysGlyGluLysGlyGlyAspGlyCysThrAlaC  726

1930 GGCCAAGTGCCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGA  1881
     ||||| ::: ||||||
 726 ysProSerLeuGlnGlyThrValThrAspMetAlaGlyLysArgProGlyGln  742

1880 CCAGACCCAGGCCTCGCGGCAGACACCATAGGCAGTGACAGACTGGCTG   1831
     ||| ||| :::|||| |||    :::
 743 ProGlyProLysGlyGluGlnGlyLeuGluGlyValGlyArgProGlyLy   759

1830 AGCTGGACAATGAGCCCATAAACAGGATGGGCGCCACCTGGACAGCAG    1781
     ||||||  ::::||||||||| |
 759 sProGlyGlnProGlyLeuProGlyValGlnGlyProProGlyLeuLysG  776

1780 GA...AGGCACTATCCAGGATGGCGAGGTCCAGGCAGA...........  1746
     |||||| ||||||
 776 lyValGlnGlyGluProGly......ProProGlyArgGlyValGlnGly  790

1745 ......................................TGCCC        1741
                                           ||
 791 ProGlnGlyGluProGlyAlaProGlyLeuProGlyIleGlnGlyLeuPr  807

1740 CGGCCCGGAA...CCACCCTGGCCT..........CGGTGGGCTCACCCAC  1703
     |||||||::: |||||| |||    :::||||
 807 oGlyProArgGlyProProGlyProThrGlyLysGlyAlaGlnGlyLys   824

1702 CACCACACGTACGGAGACATCACAGGC...................      1676
     ||||||
 824 erProGlyValLys.GlyAlaThrGlyProValGlyProProGlyAlaIase  840

1675 .......................................A           1675
                                            ||
 840 rValSerGlyProProGlyArgAspGlyGlnGlnGlyGlnThrGlyLeuA  857

1674 GAGGC...CCCGCAGAGCGCGGTGGAGGTGGAGCAGGCCACTCCCTCC   1628
     |||||  :::||||  |||  :::::::
 857 rgGlyThrProGlyLeuLysGlyProArg.GlyGluLysGlyGluProGl  873

1627 AGCACCCAGTGTCCATTAGGG................AAGG          1602
     ::|||||||
 873 yGluCysSerCysProSerGlnGlyAspLeuIlePheSerGlyMetProG  890

1601 GAGCTCCAGGCTTAGGGCCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCA  1552
     ||||||||||| :::|||||||:||
 890 lyAlaProGlyLeuTrpMetGly.SerSerTrpGlnProGly........  903

1551 CTGCTAGCACCTCCAGTGTCCCCTCCGTATTGGGCAGGAAACACCTGCTT  1502
     ||| |||  :::||||
 904 .........ProGlnGlyProProGlyIyIle..ProGlyProProGlyP  916

1501 CTCCCGGTGGTGTAGAGGGAGGCCAGTGTGTAGGGCAGGATCTGAGGGGCTG  1452
     |||||| |||  |||  :::||||
 916 roProGly.ValProGlyLeuGlnGlyVal...........ProGlyAs  928

1451 AGAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCACCACGGC........  1409
     ::|||| |||||||| :::::::|||
 928 nAsnGlyLeuProGlyGlnProGlyLeuThrAlaGluLeuGlySerLeuP  945

1408 ..........CACACTGTGGACAGGCATGTGGCACCGGCAGCCACAGG   1370
               ||| ||||||
 945 roIleGluGlnHisLeuLeuLysSerIleCysGly................  956

1369 GAAAGCTGCCACACTGCCACAAATAGACTGCTCGAGTCGCCGAATCGCTGCA  1320
```

```
2772 TTGGGGGCCAGAGCCCGAGGAGAGAAGATTCTGGCAATGATCAGCCCAA  2723
     |||||||||::
 462 ProGly.....ThrProGlyAsp......................ProGl  469

2722 TGA...CCAGTATCTCAGGGGACCTGATTGTTGGGGATCCCCCACCCT   2677
     |||  :::|||||||||| |||||
 469 yGlyProProGlyLysGlyAsp......................      477

2676 ACCCAAATATTAGCACCAACACAGAAAAGCTAGCAATGGATTCCCTTCT  2627
     |||:::|||::||||:|||
 478 ..........................LysGlySerGlyIlePro...    484

2626 ACTTTGTTAAATAAATAAGTTAAATATTTAAATGCCTGTGTCTCTGTGAT  2577
 484 ....................................               484

2576 GGCAACAGAAGGACCAACAGGCCACATCCTGATAAAAGGTAAGAGGGGGG  2527
     ::: |||||||| ||| |||||| ||||
 485 .GlyLysGluGlyProGlyGlyLys........ProGlyLysProGlyv  498

2526 TGGATCAGCAAAAGACCAGTGCTGTGGGCTGAGGGACCTGGTTCTTGTG  2477
     ||  ||||||
 498 aLysGlyGluLys.....................GlyAsp........   504

2476 TGTTGCCCCTCAGGACTCTTCCCCTACACAAATAGTCATATGTTCAAATCC  2427
                                         ||
 505 .................................Pr               505

2426 CATGGAGAGTGTTTCATCCTAGAAACTCCCATGCAAGAGCTACATTAAA  2377
     |||: :||||:||:::::|                   .
 505 ocysGluvalCysProThrLeuProGluGlyPheGlnAsnPhe......v  520

2376 CGAAGCTGCAGGTTAAGGGGCTTAGAGATGGAAACCAGGTGACTGAGTT  2327
     |||  |||::||||||||| |||
 520 aIGlyLeuProGlyLysProGlyGlyProGlyAspProVal          536

2326 TATTCA.....................GCTCCCAAAAA            2310
     :::                         ::
 537 ArgAlaAargGlyAspProGlyIleGlnProGlyIleLysGlyGluLysGlyGlyGl  553

2309 CCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCTGAGC.  2261
     :|||| |||      :::|||
 553 uProCysLeuSerCysSerSerValaIGlyAlaGlnHisLeuValSerS  570

2260 CTGGGTAATCCA.....................                 2249
     ::||||||:::
 570 erThrGlyAlaSerGlyAspValaIGlySerProGlyPheGlyLeuProGly  586

2249 ...................................              2249
                                        :::::::
 587 LeuProGlyArgAlaGlyValProGlyLeuLysGlyGluLysGlyGlyAsnPh  603

2248 .............CCTGCAGAGTCCCCG...CATTCCAGTGCATGGAGCC  2215
                  |||||| |||  :::|||
 603 eGlyGluAlaGlyProAlaGlySerProGlyProProGlyProValGlyP  620

2214 CTTCTGGCCTC..................CCTGTA              2198
     ||:::|||:::
 620 roAlaGlyIleLeuLysGlyAlaLysGlyAspValGlySerProGlyProAla  636

2197 TRAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGCAGC......  2153
     |||:::||||||:||::::
 637 .............LeuSerAsnLeuGlnAspGluGlyLysGlyAspValAr  646

2152 CCTAGAGACTGGGAGGAGAGGAGAGGGACCGCCCCCAGC            2116
     ::||||  |||||| || |||
 646 gVallAlaLeuProGlyProSerGlyGlu...LysGlyGluProGlyP   662

2115 CCCCAGCTGTGCAGCCTACGCACCTCAGCAGCACGAGGTGGCAGCAGAGAG  2066
     |||||| ::: |||||
 662 roProGlyPheGlyLeu......ProGlyLysGlnGlyIleLysGlyAlaGlyGlu  676
```

```
                                                              |||||
1210 ......ProAlaGlyLeuAspGlyLeuAspGlyLysAspGlyLysProG 1224
                                                              ||
957  ...............................AspCysValGlnGlyGlnArgAlaHi 965

570  ATGCCCAAGGACAGTGCCCAGATGAAGGGCCGGCGGG......GCCATA 527
     ::|||        |||  ||||||||:::         ::::::
1319 CCAGCCCGTCCATGACGAGAGAAGACCAGGGAGATGGCGCACTGCAGG 1270
     ::::::  ||  ::::::||||||| ||:::::
1224 lyLeuArgGly.........AspProGlyAlaGlyProProGlyLeu 1237

526  GCGTCCACGCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACAGACCA 477
     :::  ||  :|| ||  ||  |||  |||
965  sProGlyTyrLeuValGlyLeuLysGlyLysGlyAspGlnGlyIle.... 980
1238 MetGlyProProGlyPheLysGly..LysThrGlyHisProGlyLeuPro 1253

476  GGCCCAGCACTGGACCAATGCCCAGCCACCATGGTCATGAACTTCTCCT 427
     |||||      |||   ||
981  ................ProGlyVal.............ProGlyVal 985
1254 GlyProLysGlyAsp...CysGlyLysProGly.............. 1263

1219 CTCGGTGCCCGGCTCAGCTCTGGGCACGCCCTGTACAGCCCCTGCCCA 1170
     |||         :::|||     :::||| |||:::
986  LeuAspAsnCysAlaGlnCysPheLeuSerLeuGluArgProArgAlaGl 1002
426  ACCCCCA.............CTTCCAGCAGCAGAGGCG........... 402
     ||||||                 |||:::::||||:::
1264 .ProProGlySerThrGlyArgProGlyAlaGluGlyGluProGlyAlaM 1280

1169 CGAAATCCGTGTAAAACAGGCTGAAGGTCATGAGTGCCATCCAGC.TGCA 1121
     :||| |||  ||  |||||      |||  |||:
1002 uGluAlaArgGlyAspAsnSerGluGly........AspProGlyCysV 1016
401  ...........................GCAACATAGGTGATGCCTGCCG 381
                                   :::|||
1280 etGlyProGlnGlyArgProGlyProProGlyHis.........ValGly 1293

1120 CAGCTCAGCCA........................CGAAGAGCC 1101
     :::||||||||                           ::|||
1016 alGlySerProGlyLeuProGlyProProGlyLeuProGlyGlnArgGly 1032
380  CCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCAAGAGCTGG 331
     |||       |||||||||||||
1294 ProProGlyProProGlyGln.............ProGlyProAlaGl 1305

1100 GGGCAGGTGCGGGGCATGCGGCACAGCTGGTGCAGCCGGGGAAGC 1051
     :::   |||            |||   |||
1033 GluGluGlyProProGlyMetArgGlySerProGlyProProGlyProIl 1049
330  GCTTTCCGGT.... 321
     |:::|||:::
1305 yIleSerAlaValGlyLeuLysGlyAspArgGlyAlaThrGlyGluArgG 1322

1050 AGGGGCGGCCAGGT........TCCGGAAAGCCAAGCGGGGCCCGGCA 1013
     |||  ||||||                    ||  ||||||||||
1049 eGlyProProGlyPheProGlyAlaValGlySerProGlyLeuProGlyL 1066
320  .....GCCGCAGCAGGCGGCTCACCCACCACAGCCCTCGGACCATAGTGGG.. 278
     |||         ||  :::|||  ||
1322 lyLeuAlaGlyLeuProGlyGlnProGlyProProGly.HisProGlyPr 1338

1012 TGGACAGCAGTGGGGCGACAAGGAGGGGGCCGACAGCCCTTCTGCTGGCT 963
     :::::        :::|||  |||:::||| |||::|||
1066 euGlnGly..........GluArgGlyLeuThrGlyLeuThrGlyLysAsp 1078
277  ...........CCAGGCGGGTAGGGCTCAGGGGGCCGTTCAGGCACTCCAG 238
              |||||         |||:::::|||:::::||| |||
1338 oProGlyGluProGlyThrAspGlyValAlaAlaGlyLysGluGlyProProG 1355

962  CGGTGGGCCCAGCGCTGCCTCCTCAGCCA.................. 933
     :::  ||||| |||  ||||||
1079 LysGlyGluProGlyProProGlyGlnProGlyTyrProGlyAlaThrGl 1095
237  AACTGCTTCGTCCGGCTCTGCTCCAGAAGCTGCGGCCTCCTCCTTGC 188
     :::         ||:::   |||||||||:::::||| |||:::
1355 lyLysGlnGlyPheTyrGlyProProGlyProLysGlyGlyAspProGlyAla 1371

932  ...............CCAGCAGTGTGGCTGGCTCACCGCAGGTGAGGAAGATGAG 896
     :::   |||||||  |||:::||| |||
1095 yProProGlyLeuProGlyLeuLysGlyGlyGluArg.....GlyTyrThrG 1110
187  TGCCCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTCTCTGCCAGCCCTT 138
     :::  |||         |||
1371 lyLysGlnGlyPheGlyProProGlyProLysGlyGlyAspProGlyAla 1386

895  GGTGAGCAGGCCAAAGAGGCCACTCCT...................... 870
     ||  ||||||||:::::|||  |||
1110 yLysSerAlaGlyLluLysGlyGluProGlyProProGlySerGluGlyLeu 1126
1372 AlaGlyGln......LysGlyGlnAlaGlyLysGlyArgAlaGlyMe 1386

869  CCTGGGTGCCCAGGT....................AGGGGGCCAGGGCA 841
     ||||||  |||||                       ||||||||
1127 ProGlyProProGlyLeuProGlyProAlaGlyLeuProAlaArgGlyPro...... 1141
137  TGGTGCCGGTCCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGG 88
     ::|||||||:::::|||::::             :::::  |||||:::
1386 tProGlyGlyProGlyLysSerGlySerMetGly.ProValGlyProPro 1402

840  CTGGTGTGCCCAGTCAATGGCAGGCAGGCAGGAGGTAGCCAGGCCCCAAG 791
     :::||| ||  ::::::   |||:::  ::
1142 ................GlnGlyAsnSerGlyGluLysGlyA 1150
87   CACCTCAGTGGGGACAGCTCATCACTCAGATCCTCGGCCGAGGCGGCGCG 38
     ::|||   ||:::    :::|||   |||:::
1403 GlyProAlaGlyGluArgGlyHisProGlyAlaProGlyProSerGly... 1418

790  ACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCACAGTGGTCGGGT 741
     ||         |||:::|||||||  ::|||||||
1150 sp..........GlnGlyPheGlnGlyGlnProGlyPheThrGlyProThr 1163
37   GCTGTCACCCGGA 25
     ||||||||
1419 ....SerProGly 1421

740  ...............CCCGGAAGAGGTCAGAGAGCAGGGGCCTCCA...... 711
     ||||||
1164 GlySerProGlyPheProGlyPheLysValGlyLysGlySerGluGlyProGlyPr 1180

710  ...GTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCCACCACGCC 665
     :::  ||||||  ||  |||:::::|||||||    ||||||||
1180 oGlnAlaGlyLysGlySerGluGlyIleArgGlyProSerGly...LeuP 1196

664  CAGGATGATGCCAGTGCCCAGTCCACAGTCCTGGATCCGGGACAGCAGCC 615
     ||||| |||        ||||||||||:::::
1196 roGlySerProGlyProPro.GlyProProGlyIleGlnGly........ 1209

614  CTTGCTAGGCCAGCCGGC.......CCTTGGGATGAGAAAGAGGCTCAGCAGG 571
     |||||||||||              |||  ||:::::
```

Wed Sep 29 14:26:56 1999

This Page Blank (uspto)

OM of: US-09-030-606-110 to: SwissProt_37:*   out_format : pfs

Date: Sep 25, 1999   4:43 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
   -MODEL=frame+_n2p.model -DEV=xlp
   -Q=/cgn2_1/USPTO_spool/US09030606/runat_24091999_171618_29883/app_query.fasta.1
   -DB=SwissProt_37 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
   -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
   -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
   -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
   -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
   -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=score
   -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
   -NCPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-110
Query length: 3410
Database: SwissProt_37:*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
SwissProt_37:YHL1_EBV	-	398.50	340.86	3.0e-12	660	! P03181 epstein-barr virus (str
SwissProt_37:CA11_MOUSE	-	373.00	313.96	4.3e-11	1453	! P11087 mus musculus (mouse)
SwissProt_37:CA13_HUMAN	-	351.50	295.82	4.4e-10	1466	! P02461 homo sapiens (human)
SwissProt_37:CA44_HUMAN	-	348.00	291.88	6.3e-10	1690	! P53420 homo sapiens (human)
SwissProt_37:CA13_BOVIN	-	344.50	292.24	9.6e-10	1049	! P04258 bos taurus (bovine).
SwissProt_37:CA11_CHICK	-	344.00	289.57	9.8e-10	1453	! P02457 gallus gallus (chick
SwissProt_37:CA11_HUMAN	-	343.00	288.68	1.1e-09	1464	! P02452 homo sapiens (human)
SwissProt_37:CA15_HUMAN	-	337.50	282.48	1.9e-09	1838	! P20908 homo sapiens (human)
SwissProt_37:CA1F_HUMAN	-	330.00	277.12	4.4e-09	1603	! Q07092 homo sapiens (human)
SwissProt_37:CA12_MOUSE	-	326.00	274.40	6.8e-09	1459	! P28481 mus musculus (mouse)
SwissProt_37:STP_SPIOL	+	327.00	282.31	6.9e-09	525	! Q03411 spinacia oleracea (spi
SwissProt_37:CA25_HUMAN	-	324.50	272.97	8.0e-09	1496	! P05997 homo sapiens (human)
SwissProt_37:CA12_HUMAN	-	324.50	273.34	8.1e-09	1418	! P02458 homo sapiens (human)
SwissProt_37:CA54_HUMAN	+	323.00	270.88	9.3e-09	1685	! P29400 homo sapiens (human)
SwissProt_37:CA11_BOVIN	-	321.00	274.54	1.3e-08	779	! P02453 bos taurus (bovine).
SwissProt_37:CA1B_HUMAN	-	316.00	264.52	2.0e-08	1806	! P12107 homo sapiens (human)
SwissProt_37:CA17_HUMAN	-	314.50	259.88	2.2e-08	2944	! Q00388 homo sapiens (human)
SwissProt_37:CA24_ASCSU	+	312.50	261.74	2.9e-08	1763	! P27393 ascaris suum (pig rc
SwissProt_37:CA25_HUMAN	+	312.00	262.46	3.1e-08	1496	! P05997 homo sapiens (human)
SwissProt_37:CA21_CHICK	+	311.50	262.68	3.3e-08	1362	! P02467 gallus gallus (chick
SwissProt_37:CA11_RAT	-	310.00	266.32	4.2e-08	671	! P02454 rattus norvegicus (rat)
SwissProt_37:CA34_HUMAN	-	309.00	259.17	4.2e-08	1670	! Q01955 homo sapiens (human)
SwissProt_37:CA13_BOVIN	+	308.00	261.55	4.9e-08	1049	! P04258 bos taurus (bovine).
SwissProt_37:CA54_HUMAN	-	307.00	257.43	5.2e-08	1685	! P29400 homo sapiens (human)
SwissProt_37:CA1A_HUMAN	-	306.50	263.28	6.1e-08	680	! Q03692 homo sapiens (human)
SwissProt_37:CA2B_MOUSE	-	305.50	256.56	6.2e-08	1592	! Q64739 mus musculus (mouse)
SwissProt_37:CA24_HUMAN	+	304.00	254.79	7.2e-08	1712	! P08572 homo sapiens (human)
SwissProt_37:CA34_HUMAN	+	304.00	254.97	7.2e-08	1670	! Q01955 homo sapiens (human)
SwissProt_37:CA1A_HUMAN	+	304.50	261.60	7.6e-08	680	! Q03692 homo sapiens (human)
SwissProt_37:CA24_CAEEL	+	304.50	254.19	7.6e-08	1758	! P17140 caenorhabditis elega
SwissProt_37:CA2B_HUMAN	-	303.50	254.28	7.6e-08	1736	! P13942 homo sapiens (human)
SwissProt_37:CA21_MOUSE	-	301.50	254.22	9.7e-08	1373	! Q01149 mus musculus (mouse)
SwissProt_37:CA1B_MOUSE	-	300.50	251.49	1.0e-07	1804	! Q61245 mus musculus (mouse)
SwissProt_37:CA1A_MOUSE	-	299.50	257.39	1.3e-07	680	! Q05306 mus musculus (mouse).
SwissProt_37:CA1F_HUMAN	+	298.50	250.62	1.3e-07	1603	! Q07092 homo sapiens (human)
SwissProt_37:CA11_HUMAN	+	297.00	250.41	1.5e-07	1464	! P02452 homo sapiens (human)
SwissProt_37:CA13_HUMAN	+	297.00	249.98	1.6e-07	1466	! P02461 homo sapiens (human)
SwissProt_37:CA21_HUMAN	-	296.50	250.05	1.7e-07	1366	! P08123 homo sapiens (human)
SwissProt_37:CA1B_RABIT	+	297.00	254.67	1.7e-07	744	! P14282 oryctolagus cuniculus
SwissProt_37:CA18_HUMAN	+	295.50	253.41	2.0e-07	744	! P27658 homo sapiens (human)
SwissProt_37:CA44_HUMAN	+	294.00	246.47	2.1e-07	1690	! P53420 homo sapiens (human)
SwissProt_37:CA1A_CHICK	+	294.50	253.25	2.2e-07	674	! P08125 gallus gallus (chicke

(continued)

SwissProt_37:CA11_RAT	+	294.50	253.28	2.2e-07	671	! P02454 rattus norvegicus (r
SwissProt_37:CA64_HUMAN	+	293.00	245.68	2.4e-07	1678	! Q14031 homo sapiens (hum
SwissProt_37:CA24_ASCSU	-	292.50	244.92	2.5e-07	1763	! P27393 ascaris suum (pig

seq_name: SwissProt_37:YHL1_EBV

seq_documentation_block:
ID   YHL1_EBV   STANDARD;   PRT;   660 AA.
AC   P03181;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL BHLF1 PROTEIN.
OS   EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4)
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC   GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 84270667.
RA   BAER R., BANKIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA   GIBSON T.J., HATFULL G., HUDSON G.S., SATCHWELL S.C., SEGUIN C.,
RA   TUFFNELL P.S., BARRELL B.G.;
RT   "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL   NATURE 310:207-211(1984).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; V01555; -; NOT_ANNOTATED_CDS.
DR   PIR; A03742; QQBB3.
KW   HYPOTHETICAL PROTEIN; EARLY PROTEIN; REPEAT.
FT   DOMAIN   149   648   4 X 125 AA TANDEM REPEATS.
FT   REPEAT   149   273   1.
FT   REPEAT   274   398   2.
FT   REPEAT   399   523   3.
FT   REPEAT   524   648   4.
SQ   SEQUENCE   660 AA; 66244 MW; 372F08C5 CRC32;

alignment_scores:
          Quality: 398.50   Length: 860
          Ratio: 1.126   Gaps: 54
   Percent Similarity: 41.163   Percent Identity: 27.558

alignment_block:
US-09-030-606-110/rev x YHL1_EBV

Align seg 1/1 to: YHL1_EBV from: 1 to: 660

2297 TGTGTCTCAACTAGGAGCTAGCTGTTAACCCTGAGCCTGGTAATCCAC 2248
     ||!||!||||||!|||!|||| |:::||!||| |||
   5 CysGlnSerAlaArgGlyProArgThrProLeuPro........ 17

2247 CTGCAGAGACTCCCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTA 2198
                                       ::!||| ||||||||||
  18 ....HisCysPro........ProProCysL 24

2197 TA......AGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGCAG 2154
     :: !|!||||| :::   !|||||| |!|||
  24 euProGlyAlaProAspGlnThrArgArgLeuProPro.......... 37

2153 CCCTAGAGACTGGGGAGAGGAGGACGCGCCCCAGCCCCAGCTGTGC 2104
     |!:||||||||||||| ||||||| ||!|||
  38 ....GlyTrpGlyGlnArgThrAlaProThrGlnValGlyLyLe 50

2103 AGCTACGCACCTCAGCAGCACAGGTGCAGCAGAGAGCCACATTACTTT 2054
     ||! |:!:||!::: ||| !:!:||:
  50 uAlaAspAlaLaSerProAspGluLeuGlnAspGln........... 62

```
2053 GGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCATGGGCTAACAG 2004
                     ||||||    |||    ::     ::
  63 ...............AlaSerGlyLyAlaAargProGlyGlyGlyAsn 72

2003 GAGCGGGAGCTGGGAACGTGAGGCAGGCCCTCACCCCAATGTGCTG 1954
     :::       |||||||| ||||||||   :||||||
  73 ArgValGlyAlaGlyArg.....GlyArgProGlyThrProAla 85

1953 GAAGTTTTCTACGCTGAGTAGTATTGGCCAAGTCCTCTTGTCAAATACTAC 1904
                         |||||||
  86 ....................ProSerArg.................. 88

1903 CTGTGTAGCAAAGTAAATGGCAGCCAGACCCA.............. 1872
     |||||:::     |||      |||
  89 .........GlnSerArg..ArgThrGlyProAlaGluGlnAlaAspHis 101

1871 ...............GGCCTGCGGCAGACACCATATAGGCAGTGACA 1840
                  :::   |||   |||   ::::::  :
 102 AlaHisSerAsnProThrGlyLyCysSerAspProGlnArgSerProAr 118

1839 GACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGATGGGGCCACCTG 1790
     :|!:      |||||   |:::::|||
 118 gThrArgGlnAlaGlyTyr..........AlaLeuGly...GluG 129

1789 GGACAGCAGGAAGGCACTATCCAGGATGGCGAGTCCA........... 1752
     |!:::|||||   |||   ||||||||
 129 lySerAlaGlyLeu.......GlySerArgGlyProArgProHisPro 142

1751 ...............GGCAGATGCCCCGGCCCGGCGAACCACCCTG 1723
                  ::|||| ||||| ||
 143 AlaPheGlnGlnValGlnTrpSerAlaArgAsnProGlyCysProArgThrTr 159

1722 GCCTCCGGTGGCT...........CACCCACCACCACACCTACGGA 1688
     |||   :::      ||||||||||
 159 pArgArgSerGlyAlaGlnArgGlyHisProProPro.......... 172

1687 GACATCACAGGCAGAGGCGGCCCGCAGAGCGCGGGGTGGAGGTGGGAGCAGGC 1638
                          |||||||   ||||||||
 172 ...................GlyAlaGly 175

1637 CACTGCCTCCCAGCACCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTA 1588
 175 .................................................. 175

1587 GGGCCTGGCAGGAAGCTGGTCATCAGGCTGTCTCCTCACTGCTAGCACCTCC 1538
 176 .................................................. 176

 176 .............................................G 176

1537 AGTGTCCCCTCCGGTATTTGGGCAGGAACACCTGCTTCTCCCCGTGGTAGA 1488
     |||||||     ||||||||      :::|||||||
 176 lnArgProSerGlyProThrGlyLyGlyAlaProGlyAlaaAlaProGlyAlaPro 192

1487 GGGAGGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCG 1438
     :::          ||||||||:
 193 GlyThrProAlaAlaProGlyProGlyGlyLyGlyAlaAlaValProSer.. 208

 208 ................................................. 208

1437 GTGAGGGGCCGGCTGAAGCTGTCACCACGGCCACACTGTGGACAGGC.... 1392
     ||:||:||    |||   |||||||
 209 .............GlyAlaThrProHisProGluArgGlySerGlyProA 221

1391 ...............ATGTGGCACCGGCAGCCACCAGGGAAAGCTGCCCACAC 1356
                  ::   |||  ||||||
 221 laAspProProAlaAlaAlaArgLeuProGluProGlnLyProArg 237

1355 TGGCCAAATAGACTGCTCAGTGCCGAATCGCTGCAACAGCCGGTCCATG 1306
     |||:        |||||||||
 238 LeuProGlnAspLeuAlaAlaAlaGln..ArgCysProAlaGlyLyProPr 253

1305 ACCAGAGAAGA..CCAGGGAGTAGGCCGCACTGCAGGAACAGCCCCAG 1259
     :::!|||      ||||
 253 oProThrArgSerGlyAlaAlaAlaAlaGlnArgThrHisArgArgProProG 270
```

```
1258 GCTGCCCATCCGAACGCCTTCATCATAGTGTCTTCCGGCCTCGGTGTGCCCG 1209
     ||||||||   ::||:||  :::||||  |||||||   ||||||
 270 lyCysPro.............ArgSerAlaArgAsnProGlyCysPro 281

1208 GCTCAGCTCTGGCACGCCCTGGTACAGCCCCTGGTACAGCCCCGAAATCCGTG 1159
     :::   |||   |||  :::|||     |||
 282 ..ArgThrTrpArgArgSerGlyAlaGlnArgGlyHisProPro.. 296

1158 TAAAACAGCGCTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTGCAGCCAC 1109
 296 .................................................. 296

1108 GAAGAGCCGGCGCAGGGTGCGGCGGCATGCGGCAGCACGCTGGTGCAGCC 1059
     ||||||||   |||   ||||||||     ||
 297 ....ProGlyAlaGlyGlnArgProSerGlyProThrGlyLyArgPro 311

1058 GGGGAGCAGGGCGCCCAGGTTCCGGAAAGCCAACGCGGGCCGGCATGGA 1009
     :::     |||||||   ||||||||
 312 AlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyGl 328

1008 CAGC.........AGTGGGGCGACA.............AGGAGGGGG 984
     :::         |||||||||      |||
 328 yAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerGlyP 345

 983 ................CCGACAGCCCTTCTGCTG......GCTCCGGTGGGGCCCA 951
                    |||||||||    |||   |||
 345 roAlaAsProProAlaAlaAlaArgLeuProProGluArgGlnGluPro 361

 950 GCGCTGCCTCCTCAG................................CCACC 931
     :::                                       ||||||
 362 ArgLeuProGlnAspLeuAlaLeuAlaAlaGlnArgCysProAlaGlyProPr 378

 930 AGCAGTGTGGCTGCTACGCAGTGAGGAAGATGAGGGTGAGCAGGCCAAA 881
     |||   ||  |||   ||::|||
 378 oProThr.........ArgSerGlyAlaAlaAlaAlaGlnA 388

 880 GA......GGCACTCCTCCTCGGTGCCCA..........GGTAGGGGCCAGGG 843
     ||      ||||||||||||      ::||||||||
 388 rgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGly 404

 842 CACTGGTGTCCCAGTCAATGGCAGGCAGGAGGTAGCCCAGGCAGCCCCCA 793
     ||||||   |||:::
 405 ......CysProArgThrTrpArg......................... 410

 792 AGACTGATCATGAGAGGCATAGACAGGTAGGCCCTGGCGACAGTGGTCCGG 743
     |||:||||   :::   |||
 411 ...........ArgArgSerGlyAlaGlnArgGlyHisProP 421

 742 GTCCCGGAAGAGGTCAGAGA.................GCAGGGCCTCCAGTG 708
     ||||||   ||||||||      ::|||   ||::::
 421 roProGlyAlaGlyGlnArgProSerGlyProThrGlyLyArgProAla 437

 707 GAGTGAAGCACACTGGCCACAGAAGTCCAGCAGCCCACCGCCAGGATG 658
     :::||||||||||| ||:||||||||
 438 AlaProGlyAlaProGly.....ThrProAlaAlaProGlyProGlyGl 452

 657 AGCAGTGCCAGCTCCA...GGGGCTG.........GGATCCG 627
     ::::||   ::|||  |||||   ||||||
 452 yGlyAlaAlaValProSerGlyAlaThrProHisProGluArgGlySerG 469

 626 GGCACAGCAGCCCTGCTAGCCCAGCCGGCCCTTGGGATGAGAAAGAGGCTC 577
     ||   :::||:|||   |||     :::::::
 469 lyProAlaAspProProAlaAlaAlaAlaArgLeuProProGluArgGlnGlu 485

 576 AGCAGGATGCCCAAGGAC........................... 559
     ||   ::!||||    ||||||
 486 ProArgLeuProGlnAspLeuAlaAlaAlaAlaGln..ArgCysProAlaGlyPr 502

 558 .............AGTCCCCAGAATGAAGGGCCGGCGGCGGC 531
                  ::!||||    ||||||
 502 oProProThrArgSerGlyAlaAlaAlaAlaGlnArgThrHisArgArgProP 519

 530 CATAGGCTCCACGCCAGTGGTCACTGCTGAGCCTAGGACGCGGGGACACAG 481
```

```
519 ro................GlyCysProArgSer.AlaArgAs 527
              |||||||||  ::::::::::
480 ACCAGGCCCAGCACTGGACCACCATGCCAGCACCATGGTCATGAACTTCTC 431
    :||||||          |||||        |||
527 nProGly...............CysProArgThrTrpArgArgSerG 539
    |||||||                |||||||  :::::::
430 CTCTA.........CCCCACTTCCAGCAGCAGAGGCGGCACATAGG 393
    ||||||        |||:::::::
539 lyAlaGlnArgGlyHisProProGlyAlaGlyGln............ 551
    ||||||::          ||| ||
392 TGATGCTGGGGCCAAACACCACCTCCAGGCCAAAGGTTAGCAGGTTGACC 343
    ||||||:::       |||  ||| ||          ||
552 ........ArgProSerGlyProThrGlyGlyArg.............Pr 561
          |||||||||            |||
342 AGCAAGAGCTGGGCTTTCCGGTGCCGCAGCAGGCGGCTCACCCAGACCT 293
    |||  |||          |||:::|||:::|||
561 oAlaAlaProGly...............AlaProGlyThrProAlaAlaP 573
    |||||                    |||||||
292 CTGGACCAATAGTGGGCCAGGCGGGTAGGGCTCAGGGGGCCGTTCAGGCAC 243
    |||||| :::||||  ||||||            :::||
573 roGlyPro.GlyGlyGlyAlaAlaValProSerGly.......AlaTh 586
    |||||| :::||||             ||||||
242 TCCAGAACTGCTTCGTCTCCGGCTTGCTCCAGAAGCTGCGGCCTCCCTC 193
    ||||:::       |||    |||:::|||         |||:::||||||
586 rProHisProGluArg...GlySerGlyPro.........AlaAspProP 599
    |||:::                       ::
192 CTTGCTGCCCCAACTGCCTAGGAATCAGCCAGGCGGCCCCATTCTGCCAG 143
    |:::::::    ::||||||       :::   |||    |||
599 roAlaAlaAlaArgLeuProGluArg.GlnGluProArgLeuProGly 615
    ||::::        ::|||||  ||| ||:::||
142 CCCTTTGGTGCGGTCCAGCTTCTCAGCCCCATGCTCAACACCTGCTGCTG 93
    |:::||   :::|||   |||:::|  |||    ::
615 nAspLeuAlaAlaAlaGln...ArgCysProAlaGlyProProProThrA 631
    |:::||||||||||       |||:::||| |||    :::|||
92 TGGGGCACCTCAGTGGGGACGACGTCTCATCACTCAGATCCTGGC....... 49
    ||||||          ||||:::|||         |||:::|||
631 rgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCysPro 647
    :::|||||:::|||
48 CGAGGCGCGGCGGCGTCACCCGGA 25
    ||:::||||  :::||||||
648 ArgSerAlaArg...AsnProGly 654
```

seq_name: SwissProt_37:Ca11_MOUSE

seq_documentation_block:
```
ID   CA11_MOUSE      STANDARD;      PRT;   1453 AA.
AC   P11087; Q60635;
DT   01-JUL-1989 (REL. 11, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1 OR COLA1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RX   MEDLINE; 96033240.
RA   LI S.W., KHILLAN J., PROCKOP D.J.;
RT   "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT   of type I procollagen.";
RL   MATRIX BIOL. 14:593-595(1995).
RN   [2]
RP   SEQUENCE OF 518-1128 FROM N.A.
RX   MEDLINE; 8613403.
RA   FRENCH B.T., LEE W.-H., MAUL G.G.;
RT   "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
RT   collagen protein.";
RL   GENE 39:311-312(1985).
RN   [3]
RP   SEQUENCE OF 735-1130 FROM N.A.
RX   MEDLINE; 83141374.
RA   MONSON J.M., FRIEDMAN J., MCCARTHY B.J.;
RT   "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
RT   evidence for a mouse B1 element within the gene.";
RL   MOL. CELL. BIOL. 2:1362-1371(1982).
RN   [4]
RP   SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX   MEDLINE; 83157109.
RA   MONSON J.M., MCCARTHY B.J.;
RT   "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
RT   evidence for insertions or deletions in gene coding sequences.";
RL   DNA 1:59-69(1981).
RN   [5]
RP   SEQUENCE OF 1442-1453 FROM N.A.
RX   MEDLINE; 8812276.
RA   MOOSLEHNER K., HARBERS K.;
RT   "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
RT   of the 3'-untranslated region.";
RL   NUCLEIC ACIDS RES. 16:773-773(1988).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR-FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U08020; G470674; -.
DR   EMBL; X15896; G50498; -.
DR   EMBL; M14423; G192262; -.
DR   EMBL; M17491; G192264; -.
DR   EMBL; X06753; G50500; -.
DR   EMBL; K03036; G192260; JOINED.
DR   EMBL; K03029; G192260; JOINED.
DR   EMBL; K03030; G192260; JOINED.
DR   EMBL; K03031; G192260; JOINED.
DR   EMBL; K03032; G192260; JOINED.
DR   EMBL; K03033; G192260; JOINED.
DR   EMBL; K03034; G192260; JOINED.
DR   EMBL; K03035; G192260; JOINED.
DR   PIR; A23982; A23982.
DR   MGD; MGI:88467; COL1A1.
DR   PROSITE; PS01208; VWFC; 1.
KW   EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW   GLYCOPROTEIN; COLLAGEN; SIGNAL.
FT   SIGNAL        1     22
FT   PROPEP       23    151       AMINO-TERMINAL PROPEPTIDE.
FT   CHAIN       152   1207       COLLAGEN ALPHA 1(I) CHAIN.
FT   PROPEP     1208   1453       CARBOXYL-TERMINAL PROPEPTIDE.
FT   DOMAIN       29     87       VWFC.
FT   DOMAIN      162    167       TRIPLE-HELICAL REGION (N-TERMINAL).
FT   DOMAIN      168   1181       NONHELICAL REGION (N-TERMINAL).
FT   DOMAIN     1192   1207       NONHELICAL REGION (C-TERMINAL).
FT   CARBOHYD     56     56       POTENTIAL.
FT   CARBOHYD   1354   1354       POTENTIAL.
FT   SITE        734    736       CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE       1082   1084       CELL ATTACHMENT SITE (POTENTIAL).
FT   CONFLICT   1450   1450       A -> V (IN REF. 5).
SQ   SEQUENCE   1453 AA;  137944 MW;  5126DA4E CRC32;

alignment_scores:
            Quality:   375.00               Length:   1037
```

```
          Ratio:  0.856        Gaps:  63
Percent Similarity: 42.237    Percent Identity: 27.483

alignment_block:
US-09-030-606-110/rev x CA11_MOUSE

Align seg 1/1 to: CA11_MOUSE  from: 1  to: 1453


  252 lyMetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaAlaLysGlyAsp 268
      ||:::   :::                              :::  :::::||||||
 1879 CAGACCCAGGCTGCGGCAGACACCATAGGCAGTGA...CAGACTGG 1834
      |||         :::   :::   |||||              :::::||||||
  269 AlaGlyProAlaGlyPhoLysGlyGluProGlySerProGlyGluAsnGl 285

 1833 CTGAGCTGGACAATGGAGCCCATAAACAGGGATG... 1800
      |||||||   |:|||        ||:::
  285 yAlaProGlyGlnMetGlyPro...ArgGlyLeuProGlyGluArgGlyIA 301

 1799 ....GGGCCACCTGGGACAGCAGGAAGGCACTATCCAGGATGGCCAGGT 1755
          |||||||||||||||      ::  ::|||||
  301 rgProGlyProProGlyThrAlaGlyAlaArgGly.AsnAspGlyAlaVa 317

 1754 CCAGGCAGATGCCCCGCCCGGAACCACCCTGGCCTCGGTGGGCTCACCC 1705
      |||   :::      :::|||  ||| :::   |:|   :::   |||||
  317 lGlyAlaAlaGlyPro.ProGlyProThr..GlyProThrGlyProPro 332

 1704 ACCACCACGTACGGAGACATCACAGGCAGAGGCC...CCGCAGCGC 1658
      :::::      :::::||||||:::|:::||     ||||||:::||
  333 GlyPheProGlyAlaValGlyAlaLysGlyGluAlaGlyProGlnGlyAl 349

 1657 GGGTGAGGT........CCACTGCCTCCCAG 1626
      |  |::::        |||   |||||||
  349 aArgGlySerGluGlyProGlnGlyValArgGlyGluProGlyProProG 366

 1625 CACCCAGCGTCCCATAGGG...AAGGGAGCTCCAGGCTTAGGG..... 1585
      |||    :::||  ||||||         |||   |||||
  366 lyProAlaGlyIleGlyValArgGlyGluProGlyProProG 382

 1584 CCTGGCAGGAAGCTGGTCATCAGGCTCTCCTCACTGCTAGCACTCCAGT 1535
      ||||| ||  |:::|||::|:::||||
  383 ProGlyAlaLys.GlyIAlaAsnGlyAlaPro.............. 392

 1534 GTCCCCTCGGTATTTGGCAGGAACACCTGCTTCTCCCGGTGGTAGAGGG 1485
      ||||| |||   ||:::|||:::             :::|||:::||||
  393 ........Glylle...AlaGlyIAlaProGlyPheProGlyIAlaArgGly 405

 1484 AGGCCAGTGTGTAGGGCAGGATCTGCAGGGGCTGAGAAGGTGAACCCGGTG 1435
      :::|||   |||||||      ::|||    ||||||
  406 ..........ProSerGlyPro.GlnGlyProSerGlyProProGlyP 418

 1434 AGGGCGGCTGAAGCTGTCACCACGGCCACACTGTGGGACAGGCATGTGGC 1385
      |||   :::   :::   |||  |||  :::|||||||
  418 roLysGlyAsnSerGlyGluProGly...AlaProGlyAsnLysGlyAsp 433

 1384 ACCGGCAGCCACAGGGAGAAGCTGCCACACTGGCCAAATAGACTGCTCGAG 1335
      |||||   |||   ||:::|||   |||:::|:|||
  434 ThrGlyIAlaLysGlyGluProGlyIAlaThrGly.............. 444

 1334 TGCCGAATCGCTGCACCAGCCGGTCCATGACCAGGAGAAGACCAGGGAG 1285
      :::|||  |||   ||| |||:::   :::   |||   |:::|||||
  445 ..........ValIGlnGlyPro..ProGlyPyroAlaGlyGluGluGlyLy 457

 1284 ATGGCGCACTGCAGGAACA............GCCCCAGGCTGCCCAT 1250
      :   :::||||||:::           |||||
  457 sArgGlyIAlaArgGlyLysProGluProGlyLeuProGlyProProG 474

 1249 CCGAACGC.............CTTCATCATAGTGTC 1227
      :::|||||:::           :::|||
  474 lyGluArgGlyGlyProGlySerArgGlyPheProGlyIAlaAspGlyIyVal 490

 1226 TCCGGGCCTCGGTGCCCGCCTCAGCTCTGGGCACGCCCCTGGTACAGCCCC 1177
      :::||||||:::   :::||  |||:::|||   ||||||
  491 AlaGlyProLysGlyProSerGlyGlyIAlaArgGluIAlaProGlyProIAlaG 507

 1176 TCGCCCACGAAATCCGTGTAAAACAGCCTGAAGGTCATGAGTGCCATCCA 1127
      :::|||
  507 yProLysGlySerPro.................... 512

 1126 GCTCCACAGCTCAGCCACGAAGAGCCGCAGCAGGTGCGCGGGGATGCGGC 1077
```

```
 2594 TGCCTGTGT..........CTCTGATGGC.......AACAG 2569
      |||:::|||         |||        :::|||
   52 CysIleCysHisAsnGlyThrAlaValCysAspAspValGlnCysAsnGl 68

 2568 AAGGACCAACAGGCACCAATCCTGATAAAAGTAAGACGGGGGGTGGATCAG 2519
      :::   :::   |||:::|||:::|||        |||||||
   68 uGluLeuAspCysProAsnProGlnArgArg...GluGlyGly....... 81

 2518 CAAAAAGACAGTGCTGT................ 2502
      |||||
   82 ........CysCysAlaPheCysProGluGluTyrValSerProAsn 94

 2501 ..........GGGCTGAGGGGACCTGGTTCTTGTGTTGCCCCTCAG 2464
               ||:::::||||||
   95 SerGluAspProValGlyValGluGlyPro............. 103

 2463 GACTCTTCCCTACAAATAAGTCATATGTTCAAATCCCATGAGGAGTGT 2414
                                |||:::  |||
  104 .................LysGlyGlyProGlyProG 110

 2413 TTCATCCTAGAAACTCCCATGCAAGAGAGCTACATTAAACGAAGCTGCAGGT 2364
      |||||:::  |||:::
  110 lnGlyProAlaGlyGly.ProValGlyPro................ 118

 2363 TAAGGGGCTTAGAGATGGGAAACCCAGTGACTGAGTTTATTCAGCTCCCA 2314
      |||  ||||||||  |||:::::
  119 .ProGly...ArgAspGlyIleProGlyGlnProGlyLeuProGlyProP 134

 2313 AAAACCCT.....TCTCTAGGTGTGTCTCAACTGAGGAGCTAGCTGTTA 2270
      ::|||  |||||||   |||:::|||  ||||||
  134 roGlyProProGlyProProGlyProGlyLeuGlyLyAsnPheAla 150

 2269 ACC........................CTGAGCCTGGG 2256
      :::                        :::|||:::
  151 SerGlnMetSerTyrGlyTyrAspGlulysSerAlaGlyValSerValPr 167

 2255 TAATCCACCTGCGAGAGTCCCCGCCATTCCAGTGCATGGAGGCCCTTCTGGCC 2206
      :::|||  :: |||      :::|||  ||||||
  167 oGlyProMetGlyProSer............GlyProArgGlyL 178

 2205 TCCCTGTATAAGTCCAGACTGAAACCCCCTTGGAAGGCCTCCAGTCAGGC 2156
      |||||   ||||||  :::|||:::||
  178 euProGlyProProGlyIAlaProGlyPro...GlnGlyPheGln...Gly 192

 2155 AGCCCTAGAGACTGGGAGAG.................AGGAG 2130
      |||||  |||||||||                      ||
  193 ProProGlyGluProGlyProGlyGlySerGlyGlyProMetGlyProAr 209

 2129 AGGGAACGCCCCCAGCGCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGG 2080
      |||  ||| ||||||
  209 gGlyProProGlyProPro................... 215

 2079 GTGGCAGCAGAGAGCCACATTACTTGGCAGCAACAGAAACTGCGCGGCCA 2030
                      |||:::||  :::   :::   |||||
  216 ................GlyLysAsnGlyAspAspProGlyGlu 223

 2029 GCCCCGCAGCCCATGGGGCTAACAGGAGCGGGGAGCTGGACCCAGTGA 1980
      |||  :::|||   |||:::   |||  ||||||:::
  224 AlaGlyLysProGlyArgProGlyGluArgGlyProProGlyProGlnGl 240

 1979 GGCAGGGCCCTCCACCCCAATGTGCTGGAAGTTTTCTACCGCTGAGTATTTG 1930
      |||   |||   |||:::                       :::
  240 yAlaArgGlyLeuProGlyThrAlaGly..............leuProG 252

 1929 GCCAAGTCGGCTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGCGGAC 1880
```

```
512  ..............................................................                          512        362  ............................CAAAGGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCCGG  322
1076 AGCACAGCTGTGCAGCCGGCGGGGAAGCAGGGGGCCCCAGGT.............                  1038                      :::   |||||  |||||  |||||::|| ||||| :::::||
     :::  |||||   |||||||||||| |||||                                                          788  ProGlyAspArgGlyGluAlaGlyProGlyProAlaGlyPheAlaGl  804
513  GlyGluAlaGlyArgProGlyGluAlaGlyLeuProGlyAlaLysGlyLe  529
                                                                                              321  T...GCCGACAGGCCGCTCACCCACAGCCTCT...............  291
1037 .TCCGGAAGCCAAGCGGGCCCGGCATGGACAGCAGTGGGGGCACAAGGA           989                             ::: ::||| ||| |||::||| |||:::
     ::||||:|||::|||:|||||:||||| ||                                                           804  yProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspT  821
529  uThrGlySerProGlySerProGlyProAsp......GlyLysThrGlyP  544
                                                                                              290  ............................GGACCATAGTGGGCACCAGGCGG  270
988  GGGGGCCGACAGCCCTTCTGCTGGCTCGGTGGGGGCCCAGCGCTGCCT...           942                           ||||||::::::::|||
544  roProGlyProAlaGlyGlnAspGlyArgProGlyProAlaGlyProPro           560                      821  hrGlyValLysGlyAspAlaProAlaGlyProAlaGlyProAlaGl  837
     :::||  ||||||  |||
941  .....................................CCTCAGCCACCAGCAGT  925                          269  G...TAGGGCTCAGGGGGCCGTTCAGGCACTCCAGAACTGCTTCGTCTCG  223
561  GlyAlaAlaArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyTh  577                                |||  ||| |||:::  |||:::|||
                                                                                              837  yProProGlyProIleGlyAsnValGlyAlaProGly............P  850
924  GTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGGCCCAAGAGGCA  875
577  rAla.............GlyGluProGlyLysAlaGlyLGluArgGlyLyL  589                              222  GCTCTGCTCCAGAAGCTGCGGCCTCCTCCCTCCTTGCTGCCGCCAACTGCCT  173
     ::|::  ||| |||:::::||||||                                                                 ::|::|||||||||| |||:||:|||||| :::||:::
874  CTCCT.........................................            870                          850  roLysGlyProArgGlyAlaGlyAlaGlyProPro.GlyAlaThr.........  863
     |||
589  euProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAla  605                              172  AGGAATCAGCCAGGCGCCCATTCTGCCAGCCCTTTGGTGCCGGTCCAGC  123
                                                                                                   |||:::  ||| ||||||:::::||
869  ......CCTGGGTGCCCAGGT...AGGGGGCCAGGCACTGTGTCCCAG  829                                  864  .GlyPheProGlyAlaAlaGlyArgValGlyProProGlyProSerGlyA  880
     :::|||       |||||||||||||||    :::
606  GlyAlaGlnGlyAlaProGlyGlyProAlaGlyProAlaGlyGluAspGlyGlyGl  622                          122  TTCTCAGCCCATGCTCAACACCTGCTGTGTGGGGCACCTCAGT.......  79
                                                                                                   ::|||       ||||||    :::
828  TCAATGGCAGGCAGGA.........................GGTAGCCCAG  803                               880  snAlaGlyProProGlyProProGlyProValGlyLysGluGlyGlyLys  896
     :|||  |||||
622  uGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaG  639                              78   .........GGGGACACGTCTCATCACTCAGATCCTGGCCGA..........  46
                                                                                                   |||::|||:::         |||||::
802  GCAGCCCCCAAGACTGATCATCGAAGGCATAGACAGAGCTAGGCCTGGCGAC  753                              897  GlyProArgGlyGluThrGlyProAlaGlyArgProGlyuValGlyPr  913
     ||  |||  :::       ::|::|||||||   |||||||||                                                     |||:::
639  lyProProGlyGluAlaGlyLyLysProGlyGluGlnGlyValAlaProGlyAsp  655                          45   ................GGGGGCGGCTGTCACCCGGAGCC  22
                                                                                                              |||  :::  |||||||||||
752  AGTGGT...CCGGGTCCC.....GAAGAGGTCAGAGAGCAGGGGCCTCC  712                               913  oProGlyProProGlyProAlaGlyGlyGluLysGlyGluSerProGlyAla  928
656  LeuGlyAlaProGlyProSerGlyAlaAArgGlyGluArg...GlyPhePr  671
                                                                                              seq_name: SwissProt_37:CA13_HUMAN
711  AGTGGAGTGAAGCACACCTGCCACAGAAGTCCAGCAGCCCCACGCCCAG  662
     |||   :::::::|||   ::|||||||::|||||                                                    seq_documentation_block:
671  oGlyGluArgGlyValGlnGlyProProGlyProAlaGlyProArgGlyA  688                                ID  CA13_HUMAN    STANDARD;         PRT;   1466 AA.
                                                                                              AC  P02461;
661  GATGAGCAGTGCCA.............................GCTCCA  642                                 DT  21-JUL-1986 (REL. 01, CREATED)
     ::  :::::|||                                                     |||||||                DT  01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
688  snAsnGlyAlaProGlyAsnAspGlyAlaLysGlyAspThrGlyAlaPro  704                                DT  15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
                                                                                              DE  PROCOLLAGEN ALPHA 1(III) CHAIN PRECURSOR.
641  GGGGGCTGGGATCCGGGCACAGCAGCCCTGCTAGCCAGCCGGCCCTTGGG  592                                DE  COL3A1.
     ||||||              :::   |||:::                                                       GN  HOMO SAPIENS (HUMAN).
705  GlyAla......ProGlySerGlnGlyAlaProGlyLeuGlyMetPr  719                                   OS  HOMO SAPIENS (HUMAN).
                                                                                              OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
591  ATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAAGGG  542                                OC  PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
     ||||||||||||::||||| |||         :::||                                                  RN  [1]
719  oGlyGluArgGlyAlaGlyAlaGlyLeuProGly.....ProLysGlyAspA  734                              RP  SEQUENCE OF 149-1225 FROM N.A.
                                                                                              RX  MEDLINE; 8938015.
541  CCCGGCGGCGGCCATAGCGTCCACGCCAGTGGTCACTGGCTAGCCTAGGA  492                                RA  JANECZKO R.A., RAMIREZ F.;
     |||  :::       :::  |||                                                                RT  "Nucleotide and amino acid sequences of the entire human alpha 1
734  rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGly  750                                RT  (III) collagen.";
                                                                                              RL  NUCLEIC ACIDS RES. 17:6742-6742(1989).
491  GCCGGGACACAGCACCAGGCCCAGCACTGGACCAATGCCCAGCCACCATGGTC  442                              RN  [2]
751  AlaArgGlyLeuThrGlyProIle.......................  758                                   RP  SEQUENCE OF 168-398.
                                                                                              RX  MEDLINE; 77134724.
441  ATGAACTTCTCCTCTACCCCACTTCCAGCAGCAGAGGCGGCACATAGGT  392                                 RA  SEYER J.M., KANG A.H.;
     ||||||  |||||     ::|  :::  ||                                                         RT  "Covalent structure of collagen: amino acid sequence of cyanogen
759  ..............GlyProProGlyAlaGlyAlaProGlyAlaAspLysG  771                               RT  bromide peptides from the amino-terminal segment of type III collagen
391  GATGCCTGCGCCAAACACACCTCAGGC...........................  363
     ||:::   |||||||||
771  lyGluAlaGlyProSerGlyProProGlyProThrGlyAlaAArgGlyAla  787
```

RT   of human liver.";
RL   BIOCHEMISTRY 16:1158-1164(1977).
RN   [4]
RP   REVISIONS.
RA   SEYER J.M.;
RL   SUBMITTED (DEC-1977) TO THE PIR DATA BANK.
RN   [5]
RP   SEQUENCE OF 399-727.
RX   MEDLINE; 79000343.
RA   SEYER J.M., KANG A.H.;
RT   "Covalent structure of collagen: amino acid sequence of five
RT   consecutive CNBr peptides from type III collagen of human liver.";
RL   BIOCHEMISTRY 17:3404-3411(1978).
RN   [6]
RP   SEQUENCE OF 728-964.
RX   MEDLINE; 80198282.
RA   SEYER J.M., MAINARDI C., KANG A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha 1
RT   (III)-CB5 from type III collagen of human liver.";
RL   BIOCHEMISTRY 19:1583-1589(1980).
RN   [7]
RP   SEQUENCE OF 950-1466 FROM N.A.
RX   MEDLINE; 88189827.
RA   MANKOO B.S., DALGLEISH R.;
RT   "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL   NUCLEIC ACIDS RES. 16:2337-2337(1988).
RN   [8]
RP   REVISION TO 1184.
RX   MEDLINE; 89098346.
RA   MOLYNEUX K., DALGLEISH R.;
RT   "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL   NUCLEIC ACIDS RES. 16:11883-11883(1988).
RN   [9]
RP   SEQUENCE OF 1065-1466 FROM N.A.
RX   MEDLINE; 85087944.
RA   LOIDL H.R., BRINKER J.M., MAY M., PIHLAJANIEMI T., MORROW S.,
RA   ROSENBLOOM J., MYERS J.C.;
RT   "Molecular cloning and carboxyl-propeptide analysis of human type III
RT   procollagen.";
RL   NUCLEIC ACIDS RES. 12:9383-9394(1984).
RN   [10]
RP   SEQUENCE OF 965-1200.
RX   MEDLINE; 81208139.
RA   SEYER J.M., KANG A.H.;
RT   "Covalent structure of collagen: amino acid sequence of alpha
RT   1(III)-CB9 from type III collagen of human liver.";
RL   BIOCHEMISTRY 20:2621-2627(1981).
RN   [11]
RP   SEQUENCE OF 1176-1466 FROM N.A.
RX   MEDLINE; 85157600.
RA   CHU M.-L., WEIL D., DE WET W.J., BERNARD M.P., SIPPOLA M., RAMIREZ F.;
RT   "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT   (III) collagen. Partial characterization of the 3' end region of the
RT   gene.";
RL   J. BIOL. CHEM. 260:4357-4363(1985).
RN   [12]
RP   SEQUENCE OF 1161-1200 FROM N.A.
RX   MEDLINE; 86187804.
RA   MISKULIN M., DALGLEISH R., KLUVE-BECKERMAN B., RENNARD S.I.,
RA   TOLSTOSHEV P., BRANTLY M., CRYSTAL R.G.;
RT   "Human type III collagen gene expression is coordinately modulated
RT   with the type I collagen genes during fibroblast growth.";
RL   BIOCHEMISTRY 25:1408-1413(1986).
RN   [13]
RP   SEQUENCE OF 1-170 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 88303360.
RA   TOMAN D., RICCA G., DE CROMBRUGGHE B.;
RT   "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT   of human prepro alpha 1(III) collagen.";
RL   NUCLEIC ACIDS RES. 16:7201-7201(1988).
RN   [14]
RP   SEQUENCE OF 1-176 FROM N.A.

RX   MEDLINE; 89378752.
RA   BENSON-CHANDA V., SU M.W., WEIL D., CHU M.L., RAMIREZ F.;
RT   "Cloning and analysis of the 5' portion of the human type-III
RT   procollagen gene (COL3A1).";
RL   GENE 78:255-265(1989).
RN   [15]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 97255959.
RA   KUIVANIEMI H., TROMP G., PROCKOP D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   HUM. MUTAT. 9:300-315(1997).
RN   [16]
RP   VARIANT AORTIC ANEURYSM ARG-303.
RX   MEDLINE; 93293988.
RA   TROMP G., WU Y., PROCKOP D.J., MADHATHERI S.L., KLEINERT C.,
RA   EARLEY J.J., ZHUANG J., NORRGARD O., DARLING R.C., ABBOTT W.M.,
RA   COLE C.W., JAAKKOLA P., RYYNANEN M., PEARCE W.H., YAO J.S.T.,
RA   MAJAMAA K., SMULLENS S.N., GATALICA Z., FERRELL R.E., JIMENEZ S.A.,
RA   JACKSON C.E., MICHELS V.V., KAYE M., KUIVANIEMI H.;
RT   "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT   in the triple-helical domain of type III procollagen are an
RT   infrequent cause of aortic aneurysms.";
RL   J. CLIN. INVEST. 91:2539-2545(1993).
RN   [17]
RP   VARIANT THR-689.
RX   MEDLINE; 91045136.
RA   ZAFARULLAH K., KLEINERT C., TROMP G., KUIVANIEMI H., KONTUSAARI S.,
RA   WU Y., GANGULY A., PROCKOP D.J.;
RT   "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL   NUCLEIC ACIDS RES. 18:6180-6180(1990).
RN   [18]
RP   VARIANT AORTIC ANEURYSM ARG-786.
RX   MEDLINE; 91056145.
RA   KONTUSAARI S., TROMP G., KUIVANIEMI H., ROMANIC A.M., PROCKOP D.J.;
RT   "A mutation in the gene for type III procollagen (COL3A1) in a family
RT   with aortic aneurysms.";
RL   J. CLIN. INVEST. 86:1465-1473(1990).
RN   [19]
RP   VARIANT EDS-IV ARG-828.
RX   MEDLINE; 94016385.
RA   RICHARDS A., NARCISI P., LLOYD J., FERGUSON C., POPE F.M.;
RT   "The substitution of glycine 661 by arginine in type III collagen
RT   produces mutant molecules with different thermal stabilities and
RT   causes Ehlers-Danlos syndrome type IV.";
RL   J. MED. GENET. 30:690-693(1993).
RN   [20]
RP   VARIANT EDS-IV SER-957.
RX   MEDLINE; 89109135.
RA   TROMP G., KUIVANIEMI H., SHIKATA H., PROCKOP D.J.;
RT   "A single base mutation that substitutes serine for glycine 790 of
RT   the alpha 1 (III) chain of type III procollagen exposes an arginine
RT   and causes Ehlers-Danlos syndrome IV.";
RL   J. BIOL. CHEM. 264:1349-1352(1989).
RN   [21]
RP   VARIANT EDS-IV VAL-960.
RX   MEDLINE; 95268429.
RA   TROMP G., DE PAEPE A., NUYTINCK L., MADHATHERI S., KUIVANIEMI H.;
RT   "Substitution of valine for glycine 793 in type III procollagen in
RT   Ehlers-Danlos syndrome type IV.";
RL   HUM. MUTAT. 5:179-181(1995).
RN   [22]
RP   VARIANT EDS-IV GLU-1014.
RX   MEDLINE; 92316511.
RA   RICHARDS A.J., WARD P.N., NARCISI P., NICHOLLS A.C., LLOYD J.C.,
RA   POPE F.M.;
RT   "A single base mutation in the gene for type III collagen (COL3A1)
RT   converts glycine 847 to glutamic acid in a family with Ehlers-Danlos
RT   syndrome type IV. An unaffected family member is mosaic for the
RT   mutation.";
RL   HUM. GENET. 89:414-418(1992).
RN   [23]

RP   VARIANT EDS-IV ASP-1050.
RX   MEDLINE; 90037070.
RA   TROMP G., KUIVANIEMI H., STOLLE C., POPE F.M., PROCKOP D.J.;
RT   "Single base mutation in the type III procollagen gene that converts
RT   the codon for glycine 883 to aspartate in a mild variant of
RT   Ehlers-Danlos syndrome IV.";
RL   J. BIOL. CHEM. 264:19313-19317(1989).
RN   [24]
RP   VARIANT EDS-IV VAL-1077.
RX   MEDLINE; 91374480.
RA   RICHARDS A.J., LLOYD J.C., WARD P.N., DE PAEPE A., NARCISI P.,
RA   POPE F.M.;
RT   "Characterisation of a glycine to valine substitution at amino acid
RT   position 910 of the triple helical region of type III collagen in a
RT   patient with Ehlers-Danlos syndrome type IV.";
RL   J. MED. GENET. 28:458-463(1991).
RN   [25]
RP   VARIANT EDS-IV GLU-1173.
RX   MEDLINE; 93022543.
RA   JOHNSON P.H., RICHARDS A.J., POPE F.M., HOPKINSON D.A.;
RT   "A COL3A1 glycine 1006 to glutamic acid substitution in a patient

alignment_scores:
　　　　Quality:　　352.50　　　　　Length:　　1159
　　　　　Ratio:　　0.770　　　　　　Gaps:　　　67
　Percent Similarity:　39.517　　　Percent Identity:　24.935

alignment_block:
US-09-030-606-110/rev x CA13_HUMAN

Align seg 1/1　to: CA13_HUMAN　from: 1　to: 1466

```
3051 AAACAGGGTTGTGGACTGGTGGGGAAAGTTGGGGTGGGGTAGGGGAAAGTTGG 3002
     :::::||         |||||||||||||     |||     ||
 161 LysSerGlyValAlaValGlyGlyLeuAlaGlyTyrProGlyAlaGl 177
     
2537 TAAGAGGGGGTGGATCAGCAAAAGACAGTGCTGTGGCTGAGGGGACC 2488
     |||:||||||                 :::||:|||| 
 306 yGluArgGlyArgProGlyLeuProGlyAlaAlaGlyAlaArgGlyAsn. 322

3001 GGGTAGGGGAAATTTTGGGCAGTGCCTTCATCAGCCCAGTCCTAGAGAGA 2952
     |       |||                      |||::||||
 177 yProProGly................................ProProGlyPro.ProGly 186

2487 TGGTTCTTGTGTGTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATA 2438
      322 ................................................ 322

2951 GTAGAGGGG.............AGTGGAAGTGGGGGAACCAGGCTGGGCC 2914
     ::::::||||           |||||   |||:::    |
 187 ThrSerGlyHisProGlySerProGlyTyrGlnGlyProPr 203

 322 ................................................ 322

2913 AAGAGAGAGGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCCCCACCCT 2864
     |||||||||         ||||||       :||||||:::
 203 oGlyGluProGlyGlnAlaGly...ProSerGlyProProGlyProProG 219

2437 TGTTCAAATCCATGGAGGAGTGTTCATCCAGCAAAACTCCCATGCAAGA 2388
      322 ................................................ 322

2863 CTACCTTCCTTCAACACCCTAACCTTGGGTAACAGCATTGGAATTATCA 2814
     ||||||          |||||        ::
 219 ly..............................................Ala 220

 322 ................................................ 322

2813 TTTGGATGAGTAGAATTCCAAGGTCCTGGGTTAGGCATTTTGGGGGGC 2764
     :::|||    |||             ::
 221 IleGlyProSerGlyProAlaGlyLysAspGlyLysSerGlyArgProGl 237

2387 GCTACATTAAACGAAGCTGCAGGTTAAGGGGCTTAGA......GATGGGA 2344
     |||    |||     ||||||||
 323 .............................AspGlyAlaAArgGlySerAspGlyG 331

2763 CAGACCCCAGAGAGAAGAAGATTCTGGCAATGATCAGCCCAATGACCAGCT 2714
     :||||| |||::: :::           |||
 237 yArgProGlyGluArgGlyLeu..................ProGlyProProG 249

2343 AACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTG 2294
     :::|||||              |||  :::|||         |||:::
 331 lnProGly...........................ProProGlyProPro..GlyThr 340

2713 ATCTCAGGGACCTGATTGTTGTGGGGATCCCCACCCTACCCAAATATTAG 2664
     ::::::|||||||     ||| |||  :::||:||:::|
 249 lyIleLysGlyProAla.....GlyIle.ProGlyPheProGlyMetLy 263

2387 GCTACATTAAACGAAGCTGCAGGTTAAGGGGCTTAGA......GATGGGA 2344 

2663 ACACCACACAGA.........................AAAGCTAGCAATG 2638
     | |||||                        |||::||||
 263 s...GlyHisArgGlyPheAspGlyArgAsnGlyGluLysGlyGluThrG 279

2343 AACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTG 2294

2637 GATTCCCTTCTACTTTGTTAAATAAATAAGTAAATATTAAATGCCTGT 2588
     || |||::            :::|||  :::||
 279 lyAlaProGly..................LeuLysGlyGluAsnGlyLeu 289

2587 GTCTCTGTGATGGCAACAGAGAGGACCAACAGCCACCATCCTGATAAAAGG 2538
     ||||||         |||                      :::       |||
```

```
500  ProGlyGluLysGlyProAlaGlyGluArgGlyAlaProGlyProAlaGl  516
1692 ACGGAGACATCACAGGCAGGAGGCCCCGCAGAGC......GCGGGTGGAG 1650
       |||  |||   ::::|||  |||||
516  yProArgGlyAlaAlaGlyGlyGluProGlyArgGlyValProGlyGlyP  533

1649 GTGGGAGCAGGCCACTGCCTCCAGCACCCAGTGTCCATTAGGGAAGGGA 1600
       |||  || :::|||  ::   :::
533  roGlyMetArgGlyMetProGlyGlySerProGlyGlySerGlyAspGly  549

1599 GCTCCAGGTTAGGGCCTGCAGGAAGCTGGTCATCAGGCTGTCCTCACT 1550
       ||||||  ||||| |||::::: :::|||||  |||
550  LysProGly...ProProGly.SerGlnGlyGluSerGlyArgPro....  563

1549 GCTAGGCACCTCCAGTGTCCCCTCGGTATTGGCAGGAACACCTGCTTCT 1500
       ||| |||::::
564  .....................................GlyProProGlyPro  568

1499 CCCGGTGGTAGAGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAG 1450
       |||  ||||||:::|||
569  SerGlyProArgGlyGlnPro............GlyValMetGlyPhe..  580

1449 AAGGTGAACCCGGTGAGGGGCGGCTGAAGCTGTCACCACGGCCACACTGTG 1400
       ||| |||                                  :
581  .....................................ProGlyProLysGlyA  586

1399 GGACAGGCATGTGGCACCGGCCAGCCACAGGGAAAGCTGCCACACTGGCCA 1350
       ::  |||
586  snAspGly...........AlaProGlyLys...........  592

1349 AATAGACTGCTCGAGTGCCGAATCCTGACCAGCCGGTCCATGACCAGA 1300
       ::  :::  :::::|||| |||||| :::
593  .....AsnGlyLeuArgGlyAlaGlyProGlyLysProGlyProGlyProGl  605

1299 GAGAAGACCAGGGAGGAGATGGCGACTGCAGGA................ 1269
       :  |||||||::::                :::|||
605  nGlyProProGlyLysAsnGlyGluThrGlyProGlnGlyProProGlyP  622

1268 ..ACAGCCCCAGGCTGCCCATCCGAACCGCCTTCATCATAGTGTCTCCGGG 1221
       ||:::::||||                    :::  ::|||
622  roThrGlyProGlyGlyAspLys............GlyAspThrGlyGly  633

1220 CCTCGGTGCCGCCTCAGCTCTGGGCACGCCCTGGTACA............ 1182
       ||| |||   |||  :: ||||||
634  ProProGlyProGlnGlyLeuGlnGlyLeuLeuProGlyThrGlyGlyProPr  650

1182 ................................................ 1182
665  oGlyGluAsnGlyLysProGlyGluProGlyLysProGlyAspAlaGlyA  667

1181 ...........................GCCCCTCGCCCACGA 1167
       |||||| |||||
667  laProGlyAlaProGlyLysGlyLysGlyAspAlaGlyAlaProGlyGluArg  683

1166 AATCCGTGTAAAACAGCGTGAAGGTCATGCATGAGTGACGTGCCATCCAGC 1117
       ::||||  ||| ::  :: :||
684  GlyPro.....................ProGlyLeuAlaGlyAl  691

1116 TCAGCCACGAAGAGCCGGCGCAGGGTGCGGGCGCAGCAGCTG 1067
       |||::::||||||||            :::
691  aProGlyLeuArgGlyAlaGlyProGlyProGlyProGluGlyGlyLysGlyG  708

1066 GTGCAGCCGGGGAAGCAGGCAGGGCGCCCAGGTTCC.............. 1035
       ||||:::|||||||
708  lyAlaAlaGlyProProGlyProGlyAlaAlaGlyThrProGlyLeu  724

1035 ............................................. 1035
725  GlnGlyMetProGlyGluArgGlyGlyLeuGlySerProGlyProLysGl  741

1034 .........GGAAAGCCAAGCAGCCCGGCCCATGGACAGCAGTGGGGCGGACA. 993
       ||::::||:::|||||||||        :::
741  yAspLysGlyGluProGlyLeuGlyProGlyAlaAspGlyAlaValProGlyLysA  758
```

```
992  .....AGGAGGGGGCGACAGCCCTTCTCGCTGGCTCGCGGTGGGCCCAGCG  948
       ||||||||||||||
758  spGlyProArgGlyProThr..................GlyProIle  767

947  CTGCCTCCTCAGCCACCAGCAGTGGCTCTCACCCAGTGAGGAAGATG  898
       |||||  ||  ::
768  GlyProProGlyProAla.................GlyGlnPr  776

897  AGGGTGAGGCAGGCCAAAGAGGCACTTCCTCCTGGGTGCCCAGGT...AGGG  851
       ||  |||::|  ||||||  ||||||  ||||||
776  oGlyAspLysGlyGluGlyGlyAla..ProGlyLeuProGlyIleAlaG  792

850  GGCCAGGGCACTGGTGTCCCAGTCAATGGCAGCAGGAGGAGCCCAGGC  801
       ||||||||||||
792  lyPro..................ArgGlySerProGly  798

800  AGCCCCCAAGACTGATCATCGAGGCATAGACAGAGTAGGCCTGGCGACAG  751
       :::   ||    |||   ::: ||||||  :: ||||||
799  Glu..............ArgGlyGluThrGlyProProGlyProAl  809

750  TGGT...CCGGT...CCGGAAGAGTCAGAGAGCAGGGCCTCCAGTGG  707
       :||| |||||:||| ||||||  ||  ||||||
809  aGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyGlyLysGlyG  826

706  AGTGAAGCACACCTGGCCACAGAAGTCAGCAGCCCCACGCCCAGGATGA  657
       ||   :::||||||||:::::||||  ::||| |||||
826  luArgGlyAlaProGlyGluLysGlyGluGlyGlyPro...ProGlyVal  841

656  GCAGTGCCAGCTCCAGGG.........GCCTGGGATCCGGGCACAGCAGC  616
       ||| ||| |||||  ||||||
842  AlaGlyProProGlyGlySerGlyProAlaGlyProGlyProGlnGl  858

615  CCTGCTAGCCAGCCGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCC  566
       :::||| |||||| || ||  :: |||||||
858  yValLysGly.................GluArgGlySerProGlyGlyP  869

565  CAAGGACAGTGCCCAGATGAAGGGCCGGCGGCCATAGCGTCCACGCC  516
       ||||
869  ro.............................GlyAla  871

515  AGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCCAGCACT  466
       |||||||  |||||||||
872  AlaGlyPhe........ProGlyAlaArgGlyLeuProGlyPro......  883

465  GGACCAATGCCAGCACCATGTCATGAACTTCTCCTCTACCCCCACTTC  416
       ||  ||  :::||  ||  ::  :|||
884  ..................ProGlySerAsnGlyAsnProGlyProProGlyP  895

415  CAGCAGCAGAGGCGGCACATAGGTGATGCTGCGCAAACACACCTCCA  366
       ||::::::::::                        |||
895  roSerGlySer....................................Pro  899

365  GGCCAAAGGTTAGCAGGTTGACAAGAGCTGGCTTTCCGTGCCGCGC  316
       ||::|  |||  :::  ::||||||   :::||||||
900  GlyLys.......AspGlyProProGlyProAlaGlyAsnThrGlyAlaPr  914

315  AGCAGGCGGCTCACCCACACCTCTGGACCATAGTGGG...........  278
       |||||||| :::||||||||  ::  :||
914  o........GlySerProGlyValSerGlyPro.LysGlyAspAlaGlyGln  928

277  CCAGGCGGGTAGGGCTGCAGGGGCCGTTCAGGGCACTCCAGAACTGCTTCG  228
       ||||||| |||||  :::|||  ||
929  ProGlyGluLysGlySerProGlyAlaGlyGlyProPro..........  941

227  TCTCGGCTCTGCTCCA.............................G  211
       ::|||||                               |
942  ........GlyAlaProGlyProLeuGlyLeuGlyIleAlaGlyIleThrGlyAlaA  956

210  AAGCTGCGGCCTCTCCTTGCTGCCGCCAACTGCTCCTAGGAATCAGCCA  161
       ::||  ||:::||||||||||| ||:::      :::
956  rgGlyLeuAlaGlyProGlyMetProGly...ProArgGlySerPro  971
```

```
160 GGCGCCCATTCTGCCAGCCCCTTTGGTGCCGGTGCCAGCTTCTCAG..... 116
        ||| :::              :::              ||| :::::::
972 GlyProGlnGlyValLysGlyGluSerGlyLysProGlyAlaAsnGlyLe 988

115 ................CCCATGCTCAACACCTGCTGCTCTG 91
                          |||::: ::: :::  |||::: :::
988 uSerGlyGluArgGlyProProGlyGln.GlyLeuProGlyLeuAla 1004

 90 GGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTGGC......... 49
      |||  ::|||:::   :::::::|||||:::::::::|||||||
1005 Gly...ThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGl 1020

 48 .....................................CGAGGCGCGC 39
                                            |||||
1020 yLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluA 1037

 38 GGCTGTCACCCGGAGCC 22
     :: |||||||||||||
1037 snGlySerProGlyAla 1042
```

seq_name: SwissProt_37:C44A4_HUMAN

seq_documentation_block:
ID   C44A4_HUMAN   STANDARD;   PRT;   1690 AA.
AC   P53420;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   COLLAGEN ALPHA 4 (IV) CHAIN PRECURSOR.
GN   COL4A4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 95014445.
RA   LEINONEN A., MARIYAMA M., MOCHIZUKI T., TRYGGVASON K., REEDERS S.T.;
RT   "Complete primary structure of the human type IV collagen alpha 4(IV)
RT   chain. Comparison with structure and expression of the other alpha
RT   (IV) chains.";
RL   J. BIOL. CHEM. 269:26172-26177(1994).
RN   [2]
RP   SEQUENCE OF 1219-1690 FROM N.A.
RC   TISSUE=EYE;
RX   MEDLINE; 93374047.
RA   SUGIMOTO M., OOHASHI T., YOSHIOKA H., MATSUO N., NINOMIYA Y.;
RT   "cDNA isolation and partial gene structure of the human alpha 4(IV)
RT   collagen chain.";
RL   FEBS LETT. 330:112-128(1993).
RN   [3]
RP   SEQUENCE OF 1407-1507 FROM N.A.
RX   MEDLINE; 93054473.
RA   KAMAGATA Y., MATTEI M.-G., NINOMIYA Y.;
RT   "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT   alpha 4 chain of basement membrane collagen type IV and assignment of
RT   the gene to the distal long arm of human chromosome 2.";
RL   J. BIOL. CHEM. 267:23753-23758(1992).
RN   [4]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 97338662.
RA   LEMMINK H.H., SCHROEDER C.H., MONNENS L.A.H., SMEETS H.J.M.;
RT   "The clinical spectrum of type IV collagen mutations.";
RL   HUM. MUTAT. 9:477-499(1997).
RN   [5]
RP   VARIANT AS SER-1201.
RX   MEDLINE; 95078927.
RA   MOCHIZUKI T., LEMMINK H.H., MARIYAMA M., ANTIGNAC C., GUBLER M.-C.,
RA   PIRSON Y., VERELLEN-DUMOULIN C., CHAN B., SCHROEDER C.H.,
RA   SMEETS H.J., REEDERS S.T.;
RT   "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
RT   collagen genes in autosomal recessive Alport syndrome.";
RL   NAT. GENET. 8:77-82(1994).
RN   [6]
RP   VARIANT BFH GLU-897.
RX   MEDLINE; 96379660.
RA   LEMMINK H.H., NILLESEN W.N., MOCHIZUKI T., SCHRODER C.H.,
RA   BRUNNER H.G., VAN OOST B.A., MONNENS L.A.H., SMEETS H.J.M.;
RT   "Benign familial hematuria due to mutation of the type IV collagen
RT   alpha4 gene.";
RL   J. CLIN. INVEST. 98:1114-1118(1996).
CC   -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC   GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC   MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC   NIDOGEN.
CC   -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC   ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC   WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC   -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC   COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC   COCHLEA, LUNG AND BRAIN.
CC   -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.
CC   -!- DISEASE: DEFECTS IN COL4A4 ARE ASSOCIATED WITH THE TYPE II
CC   AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
CC   GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
CC   HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
CC   MALES AND FEMALES.
CC   -!- DISEASE: DEFECTS IN COL4A4 ARE ASSOCIATED WITH FAMILIAL BENIGN
CC   HEMATURIA (FBH) OR THIN BASEMENT MEMBRANE DISEASE. FBH IS
CC   CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY
CC   DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN
CC   AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS
CC   NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBH AND AS CAN BE
CC   DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT
CC   HEMATURIA AND THIN GBM AT THAT AGE.
CC   -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X81053; G574806; -.
DR   EMBL; D17391; G457161; -.
DR   MIM; 120131; -.
DR   MIM; 141200; -.
DR   MIM; 203780; -.
KW   EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; BASEMENT MEMBRANE; REPEAT;
KW   HYDROXYLATION; COLLAGEN; GLYCOPROTEIN; SIGNAL; DISEASE MUTATION.
FT   SIGNAL   1   38   POTENTIAL.
FT   CHAIN   39   1690   COLLAGEN ALPHA 4(IV) CHAIN.
FT   DOMAIN   39   64   7S DOMAIN.
FT   DOMAIN   65   1459   TRIPLE-HELICAL REGION.
FT   DOMAIN   1460   1690   NONHELICAL REGION (NC1).
FT   SITE   94   96   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   145   147   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   189   191   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   310   312   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   724   726   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   785   787   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   989   991   CELL ATTACHMENT SITE (POTENTIAL).
FT   SITE   1206   1206   CLEAVAGE (BY COLLAGENASE).
```

```
FT   SITE            1212   1214                (BY SIMILARITY).
FT   SITE                                       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID        1480   1569                OR 1566 (BY SIMILARITY).
FT   DISULFID        1513   1531                OR 1569 (BY SIMILARITY).
FT   DISULFID        1525   1531                BY SIMILARITY.
FT   DISULFID        1588   1686                OR 1683 (BY SIMILARITY).
FT   DISULFID        1622   1683                OR 1686 (BY SIMILARITY).
FT   DISULFID        1634   1641                BY SIMILARITY.
FT   CARBOHYD        142    142                 POTENTIAL.
FT   CARBOHYD        669    669                 POTENTIAL.
FT   VARIANT         897    897                 G -> E (IN FBH).
FT   VARIANT         1201   1201                G -> S (IN AS).
FT   CONFLICT        1659   1660                LQ -> FE (IN REF. 2).
SQ   SEQUENCE        1690   AA;   164095 MW;   2E7EDC5B CRC32;
```

alignment_scores:
```
        Quality:   348.50              Length:   1439
        Ratio:     0.681               Gaps:     83
Percent Similarity:  35.580     Percent Identity:  23.628
```

alignment_block:
us-09-030-606-110/rev x CA44_HUMAN

Align seg 1/1 to: CA44_HUMAN from: 1 to: 1690

```
3029 GGGAAAGTTGGGGTAGGGAAGTTGGGGGTAGGGAAATTTTGGGC..  2982
     |||  |||||||||||| ||||||  ||| |||||||  ||||||
 149 GlyPheProGlyGlyArgGlyAlaLeuGlyProGlyLeuGlyHi  165
```

```
2981 .................................AGTGCCTTCATC..........  2970
                                    ||||||||||
 165 sProGlyGlyGluLysGlyGlyAsnSerValPheIleLeuGlyAlaV  182
```

```
2969 .................................AGCCCAGTCCTAGAGAGTAGAG  2946
                                    :::||| |||
 182 alLysGlyIleGlnGlyAspArgGlyAspProGlyLeuProGlyLeuPro  198
```

```
2945 GGGAGTGGGAAGTGGGGGGGGAACCAGGCTGGGCCAAGA..........GA  2908
     ||||||  :::::||||:||:  ||||||||                   ||
 199 GlySerTrpGlyAlaGlyGlyProAlaGlyProThrGlyTyrProGlyGl  215
```

```
2907 AGAGGGGTGGTGGTAGGGAAGCCGTTGAGACCTGAAGCCCCA..........  2868
     ||| | :::||| |||||| |||||||||||
 215 uProGlyLeuValGlyProProGlyGlnProGlyArgProGlyLeuLysG  232
```

```
2868 .....................................................  2868
                                  :::|||  |||
 232 lyAsnProGlyValGlyValLysGlyValGlnMetGlyAspProGlyGluVal  248
```

```
2867 .................CCCTCTACCTTCCTTCAACACCCTAACCT  2839
                    ||||||||: :::||:|||:::||:
 249 GlyGlnGlnGlySerProGlyThrLeuLeuValGluProProAspPh  265
```

```
2838 T.........GGTAACAGCATTGGAATTATCATTGGGATGATGAGTA  2801
     :          ||| |||::: :: |||::::| ||||||| |||
 265 eCysLeuTyrLysGlyGluLysGlyIleLysGlyIleIleProGlyMet.Val  281
```

```
2800 GAATTTCCA..............................................  2792
     :::||
 282 GlyLeuProGlyProProGlyArgLysGlyGlyGluSerGlyIleIleGlyAlaLy  298
```

```
2791 ...AGGTCCTGGGTTAGGCATTTGGGGGCCA.........GACCCCAGGA  2752
     :::::||: ::     ||| ||||||       ||||||
 298 sGlyGluLysGlyIleProGlyPheProGlyArgGlyAspProGlyLyS  315
```

```
2751 GAAGAAGAATTCTGGCAATGACAGCCAATCAGCCTATCCAGGGAC  2702
     ||                 ||| ::::||::::|||:||||||||
 315 erTyrGly..........SerProGlyPheProGlyLeuLysGlyGlyGlu  327
```

```
2701 CTG...ATTGTTGGGGATCCC..................................  2684
     |||  ::|||||||||||||
 328 LeuGlyLeuValGlyAsppProGlyLeuPheGlyLeuLeuIleGlyProLysGl  344
```

Right column:

```
2683 ...................................CCACCC.........TACC  2674
                                      |||||
 344 yAspProGlyAsnArgGlyHisProGlyProProGlyValLeuValThrP  361
```

```
2673 CAAATATTAGCACACCAACACAGAAAAGTAGCAATGCAATGATTCCTTCTACT  2624
     ||       ::::      :::::  |||||||||||:|
 361 roProLeuProLeuGlyProGlyAspProGlyPheProGlyArg  377
```

```
2623 ..................................TTGTTAAA  2616
                                     |||||||
 378 TyrGlyGluThrGlyAspValGlyProProGlyProGlyLeuLeuGl  394
```

```
2615 TAAATAAGTAAATATTAAATGCCTGTGTCTCTGTGATGGCAACAGAAG  2556
     :::                ||||||||:::::::::::::::::
 394 yArgProGlyGlu.......AlaCysAlaGlyMetIleGlyProProGlyProG  408
```

```
2565 GACCAACAGGCCAC.........ATCCTGATAAAAGTAAGAGGGGGGTG  2525
     |||||  |||         ::: ||||| |||:::  ||
 408 lyProGlnGlyPheProGlyLeuProGlyLeuProGlyGluAlaGlyIle  424
```

```
2524 GATCAGACAAAAGACAGTGCTGTGGCTGAGGGGACCTGGTTCTTGTGTG  2475
     ||||::: |||||||||| |
 425 ProGlyArgProAspSerAlaProGly.........  433
```

```
2474 TTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAATCCCA  2425
     :::       ||||||||||||
 433 .......................  433
```

```
2424 TGGAGGAGTGTTTCATCCTAGAAACTCCCATGCAAGAGCTACATTAAACG  2375
 433 .......................  433
```

```
2374 AAGCTGCAGTTAAGGGGTTAGAGATGGAAACCAGGTGACTGAGTTTA  2325
     ||||||||||||||
 434 .........LysProGlyLysProGly.........  439
```

```
2324 TTCAGCTCCCAAAAACCCTTCTCTA...GGTGTGTCTCAACTA...GGAG  2281
     ::::||     |||:::||      |||     ||| |||
 440 ....SerProGlyLeuProGlyAlaProGlyLeuGlnGlyLeuProGlyLys  455
```

```
2280 GCTAGCTGTTAACCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCGCAT  2231
     |||::: :::::||||||||
 455 erSerValIleTyrCysSerValGlyAsnProGlyProGlnGlyIleLys  471
```

```
2230 TCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAGACTGAAAC  2181
     ::: :::||||:::||||||||    |||
 472 GlyLysValGlyProProGlyGlyArgGly......ProLysGlyGluLy  486
```

```
2180 CCCCTTGGAGGCCCTCAGTCAGCGAGCCCTAGAGACTGGGAGAGAGGA  2131
     ||||||||         ||| :::::||
 486 sGlyAsnGluGlyLeuCysGlyAlaCysGluProGlyPro..........M  499
```

```
2130 GAGGGACGCCCCCAGCCGTGCAGCTAGCCACCTCAGCAGCACAG  2081
     ||| |||  ||||||||||:::||     |||
 499 etGlyProProGlyProProGlyLeuProGlyArg...GlnGlySerLys  514
```

```
2080 GGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGGCGGCC  2031
     |||  ||||||||||||||:::|||||  ::::|||||
 515 GlyAspLeuProGlyLeuTrpLeuGlyThrLysGlyAspProGlyPr  531
```

```
2030 A.....GCCCGGCAGCCC.................CATGGGCTAAC.  2006
     ||:::  |||               |||||||||||::::
 531 oProGlyAlaGluGluProGlyLeuProGlyLysHisGlyAlaSerG  548
```

```
2006 ...................................................  2006
 548 lyProProGlyAsnLysGlyAlaLysGlyAspMetValValSerArgVal  564
```

```
2005 ..........AGAGCGGGAGCTGGGACCCAGTGAGGCAGGCCCTCCA....  1967
     |||||||| |||:::::     |||||||
 565 LysGlyHisLysGlyGlyGluArgGlyProAsp......GlyProProGlyPh  579
```

```
1966 .CCCAATGCTGCTGGAAGTTTCTACGCTGGAGTATTGGCCAAGTCGCTC 1918          863 |||   ||||||||:::                              |||
 579 |||  |||||||   |||   |||:::||||                                      roGlyProAlaGlyMetLys...............GlyLeuProGlyLeu 874
      eProGlyGlnProGlySerHisGlyArg.....AspGlyHisAlaGlyL 593
                                                                        1418 TCACCACGGCCACACTGTGGGACAGCCATGTGGCACCGGCACCCACAGGG 1369
1917 TTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCA......... 1878          875 ||||||      :::                             HisGlyProPro....... 884
 594 ..............................GluLysGlyAspProGlyProPro 601        ProGlyArgProGlyAla................

                                                                        1368 AAAGCTGCCACACTGGCAAATAGACTGCTCAGTGCGAATCGCTGCCAC 1319
1877 ......................GACCCAGGCCTGCGGCAGACACCATATAGGCA 1846     885 ||||||   |||:::     :::       :::||:||||
 602 GlyAspHisGluAspAlaThrProGlyGlyLyLysGlyPheProGlyProLe 618          .GlyLeuProGlyIleProGlyProPheGlyAspAspGlyLeuProGlyP 901

1845 GTGACAGACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGATGGGGC 1796          1318 CAGCCGGTCCA...TGACCAGAGAAGAACCAGGAGATGCGCCACTGCA 1272
 618 |||  ||||| |||:||||  |||:|| |||:|||                              901 |||:|||||                    ||||||
      uGlyProProGlyLysAlaGlyProValGlyPro...ProGlyLeuGlyP 634           roProGlyProLysGlyProArgGlyLeuProGly............... 912

1795 CACCTGGGACAGCAGGAGGCAC.............TATCCA 1767                   1271 GGAACAGCCCAGGCTGCCCATCCGAACGCCTTCATCATAGTGTCTCCGG 1222
 634 ||||||     |||:::::                  ::|||                        913 .......PheProGlyPheProGlyGluArgGlyLyLysProGlyAlaGluGl 927
      heProGlyProProGlyGluArgGlyHisProGlyValProGlyHisPro 650

1766 GGATGGCGAGGTCAGGCAGCAGATGCCCCGGCCCGAA............. 1731        1221 GCCTCGTGCCCGGCTCAGCTCTGGGCACGCGCCCTGTACAGCCCCTCGCC 1172
 651 |||  ||||||||| |||:::                                            927 |||:::                  :::|||||||
      GlyValArgGlyLeuLysGlyLeuLysGlyGlnLysGlyAspThrIleSe 667            Y......CysProGlyAlaLys...GlyGluProGlyGluLysGlyMetS 941

1730 .......................................C 1730                  1171 CACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGCTGC 1122
 667 rCysAsnValThrTyrProGlyArgHisGlyProProGlyPheAspGlyP 684            941 ||| 
                                                                             erGlyLeuPro....................................... 944

1729 CACCCTGGCTCGGTG................................... 1714        1121 ACAGCTCAGCCACGAGACGCCGGCGCAGGGTGCGGGGCATGCGGC..... 1077
 684 |||| |||:::                                                      945 ||||||  :::||||   :::|||||||
      roProGlyProLysGlyPheProGlyProGlnGlyAlaProGlyLeuSer 700           ......GlyAspArgGlyLeuArgGlyAlaLysGlyAlaIleGlyProPr 959

1713 ............GGCTCACCCACCACCACCACGTACGGAGGACAATC 1682          1076 .........................AGCACAGCTGGTGCAGCCGCGGGG 1055
 701 GlySerAspGlyHisLysGlyArgProGlyThrProGlyThr........ 714           959 oGlyAspGluGlyGluMetAlaIleIleSerGlnLysGlyThrProGlyG 976

1681 ACAGGCAGAGGCCCCGCAGACGCGGGT....................... 1654         1054 AAGCAGGGGCGCCCAGGTTCC............GGAAAGCCA 1026
 715 |||| |||   |||:::                                                976 ||  ||| ||||||:::
      .AlaGluIleProGlyProProGlyPheAspArgGlyAspMetGlyAspP 730          luProGlyProProGlyAspAspGlyPheProGlyGluArgGlyAspLys 992

1653 ........GGAGGT...GGGACAGGCCACTG...CCTCCAGCACCC 1621          1025 AGCGGGGCCCGGCATGAGACAGCAGTGGGGCGACAAGGAGGGGGCCGACAGC 976
 730 ||||||   |||:::  ||||| |||:::                                    993 :::|||||||||                   ||||||||
      roGlyPheGlyGlyGluLysGlySerProValGlyLeuProProGlyPro 746           GlyThrProGlyMetGlnGly.........ArgArgGlyGlu....... 1003

1620 ACGTGTCCATTAGGG.........AAGGGAGCTCCAGGC........ 1591         975 CCTTCTGCTGCTCGGTGG............................. 957
 747 ||||| |||:::              ||||||                                 1004 ||:::|||||                              |||||||
      ProGlySerProGlyValAlaAsnGlyGlnLysGlyIleProGlyAspProAl 763        ........LeuGlyArgTyrGlyTyrProGlyProGlyPheHisArgGlyGluProG 1018

1590 ...........TTAGGG....................CCTG 1581              956 .......................GGCCCAGGCGCTGCCTCCTCCAGCCACCAGCAGTG 924
                                                                        1018 lyGluLysGlyGlnProGlyProGlyProProGlyProProGlyLySer 1034
1580 GCAGGAAGCTGGTCATCAGGCTGTCTCCTCACTGTCTAGCACCTCCAGTGTCC 1531     1531
 763 aPheGlyHisLeuGlyProProGlyLysArgGlyLeuSerGlyValProG 780         923 TGGCTGCTACGCAGG...TGAGGAAGATGAGGGTGAGCAGGCCAAAGAGG 877

1530 CCTCGGTATT....................... 1521                         876 CACTCCT.............CCTGGGTGCCCAGGT..................A 854
 780 lyIleLys.GlyProAspAspGlyCysProGlyAlaGluGlyPr 796               1051 yGluProGlySerProGlyProProGlyPheSerGlyIleAspGlyAlaA 1068

1521 ..........................AAGGGAGCTCCAGGC... 1521             853 GGGGGCCA...............GGGCACTGG................. 837
 796 oAlaGlyIleProGlyPheLeuGlyLeuLysGlyLeuLysProGlyArgGluG 813

1520 ..................TGGGCAGG 1513                               836 ...........................TGTCCCAGTCAATGGCAGGC 817
 813 lyHisAlaGlyPheProGlyValProGlyProProGlyHisSerCysGlu 829        1068 rgGlyProLysGlyAsnLysGlyAspProAlaSerHisPheGlyProPro 1084

1512 AACACCTGCTTCTCCCGGTGGTAGAGGAGGCCA......GTGTCTAGGG 1469       1085 GlyProLysGlyGluProGlyGlySerProGlyLyCysProGlyHisGlyAl 1101
 830 ArgGlyAlaProGlyIleProGlyLeuProGlyGlnProGlyLeuProGlyTyrProGl 846

1468 CAGGATCTGCAGGGCTGAGAGGTGAACCCGGTGAGGGGCGGCTGAAGCTG 1419      816 AGGAGGTAGCCCAGGCAGCCCCAGACGCCCAAGACTGATCATGAAGGCATAGACAGA 767
                                                                        1101 aSerGlyGluGlnGlyLeuProGlyGlnGlyLeuProArgGlySerProG 1118

                                                                        766 GTAGGCCTGGCCACAGTGGT............................CCGGGT... 741
```

```
1118 lyArgProGlyProGlySerSerGlyProProGlyCysProGlyProGlyAsp 1134
            ||||||
 740 ........CCCGGA..AGAGGTCAGAGAGCAGGGGCCTCCAGTGGAGTG    703
            |||||| |||||||||
1135 HisGlyMetProGlyLeuArgGlyGln...........ProGlyGluMe    1147
                                               ||| |||
 702 AAGCACACCTGCCACAGA..............AGTCAGCAGCCCCA...C    668
        ::: |||||| |||            ::|||:::::  |||
1147 tGlyAspProGlyProArgGlyLeuGlnGlyAspProGlyIleProGlyP   1164
            |||||||
 667 GCCCAGGATGAGCAGTGCCA.............................   648
        |||||        |||
1164 roProGlyIleLysGlyProSerGlySerProGlyLeuAsnGlyLeuHis   1180
        ||           |||
 647 ................GCTCCAGGGGCCTGGGATCC               628
                            |||  ||| |||
1181 GlyLeuLysGlyGlnLysGlyThrLysGlyAlaSerGlyLeuHisAspVa   1197
            ::|||||         |||    ||||||||
1181 GGGCACAGCAGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAGAGGCT   578
        ||| :::  :::::|||          ||||||||||:
1197 lGlyProProGlyProValGlyIleProGlyLeuLysGlyGluArgGlyA   1214
            ||||||||          |||||||:|
 577 CAGCAGGATGCCCAAGGACAGTG.......CCCAGATGAAGGGCCGGC    537
        ::: |||  |||                  ||||||
1214 spProGlySerProGlyIleSerProProGlyProArgGlyLysLysGly   1230
               |||         |||
 536 GGCGGCCATAGCGTCCACGCCAGTGTCACTGGCTGAGCCTAGGAGCCGGG   487
        ::::||||||              |||||
1231 ProProGlyProProGlySerSerGlyProProGlyProAlaGlyAlaTh   1247
              ||  ||||||||
 486 ACACAGA..............CCAGGCCCAGCACTGGACCAAT        458
        ||         |||||||||    |||||||||
1247 rGlyArgGlyAlaProLysAspIleProAspProGlyProGlyAspGlnG   1264
                ||||||||||
 457 GCCCA........................                      453
        || |
1264 lyProProGlyProAspGlyProArgGlyAlaProGlyProProGlyLeu   1280
              ::|||||||
 452 ..............GCACCATGGTCATGAACTTCTCC             430
                  ||| :::   ||
1281 ProGlySerValAspLeuLeuArgGlyLeuProGlyAspCysGlyLeuPr   1297
            ||||||||
 429 TCTACCCCACTTCCAGCAGCAGAGGCGGCACATAGGTGATGCCTGCGGC   380
        |        ||    |||    |:: ||      |||||||||
1297 oGlyProProGlyProProGly...........................P  1305
            |||||||||
 379 CAAACACAACCTCCAGGCCAAGG..........TTAGCA           351
        || ||||||||               ::
1305 roProGlyProProGlyTyrLysGlyPheProGlyCysAspGlyLysAsp  1321
            |||||||
1338 heProGlyProProGlyGluLysGlyLeuProGlyProProGlyArgLys 1354
        ||| |||         ||||||||
 350 GGTTGACCAGCAAGAGCTGGGCTTTCCGTGCCGCAG...CAGGCGGCTC 304
        ||| ::|||:::  :::||||||||   ||||||
1322 Gly..GlnLnLysGlyProMetGlyProPheProGlyLeuProHisGlyP 1338
              ||:::::
 303 ACCCAACAGCCTCTGGACCATAGTGG...CCAGGCGGGTAGGGCTCAGGG 257
        ||| |||  ||||||||   :::::||||
1338 heProGlyProProGlyGluLysGlyLeuProGlyProProGlyArgLys 1354
1355 GlyProThrGlyLeuPro.......................GlyProArgGl 1364
        ||| ::|||  |||                        :::||| |||
 256 GGCCGTTCAGGCACTCCAGAACTCCTTCGTCTCCGGCTCCTGCTCCAGAAGC 207
        ||   :::|||  |||                     ::|||:::::
 206 TGCGGGCCTCTCCTTGC..........TGCCGCCAACTGCCTAGGA 169
        ||:::||||||         ::|||:::||| |||
1364 yGluProGlyProAlaAspValAspAspCysProArgIleProGly 1380
1381 LeuProGlyAlaProGlyMetArgGlyPro.................... 1390
        ||| |||  |||
 168 ATCAGCCAGGCGCGCCATTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCT 119
        ::: ||| |||          ||||||
1391 ...........GluGlyAlaMetGly...LeuProGly....... 1398
        ::|||:::|||   :::|||
 118 CAGGCCCATGTCAACACCTGCTGTGGGGCACTCAGTGGGACACGT 69
```

```
  68 CTCATCACTCAGATCCTGGCCCAGGCGCCGGCTGTCACCCGGAGCCAGC 19
     :::::::  |||||| ||||||| :::||||||  :::
1399 ..MetArgGlyProProGlyProGlyCysLysGlyGluProGlyLeuAsp 1414
1399 ..MetArgGlyProProGlyProGlyCysLysGlyGluProGlyLeuAsp 1414

  18 GGCTGCAGG 10
     ::: |||
1415 GlyArgArg 1417
```

seq_name: SwissProt_37:CA13_BOVIN

seq_documentation_block:
```
ID   CA13_BOVIN    STANDARD;           PRT;   1049 AA.
AC   P04258;
DT   20-MAR-1987 (REL. 04, CREATED)
DT   20-MAR-1987 (REL. 04, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   COLLAGEN ALPHA 1(III) CHAIN.
GN   COL3A1.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE OF 1-242.
RX   MEDLINE; 80026026.
RA   FIETZEK P.P., ALLMANN H., RAUTERBERG J., HENKEL W., WACHTER E.,
RA   KUHN K.;
RT   "The covalent structure of calf skin type III collagen. I. The amino
RT   acid sequence of the amino terminal region of the alpha 1(III) chain
RT   (positions 1-222).";
RL   HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:809-820(1979).
RN   [2]
RP   SEQUENCE OF 243-422.
RX   MEDLINE; 80026027.
RA   DEWES H., FIETZEK P.P., KUHN K.;
RT   "The covalent structure of calf skin type III collagen. II. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2
RT   (positions 223-402).";
RL   HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:821-832(1979).
RN   [3]
RP   SEQUENCE OF 423-571.
RX   MEDLINE; 80026028.
RA   BENTZ H., FIETZEK P.P., KUHN K.;
RT   "The covalent structure of calf skin type III collagen. III. The
RT   amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4
RT   (positions 403-551).";
RL   HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:833-840(1979).
RN   [4]
RP   SEQUENCE OF 572-808.
RX   MEDLINE; 80026029.
RA   LANG H., GLANVILLE R.W., FIETZEK P.P., KUHN K.;
RT   "The covalent structure of calf skin type III collagen. IV. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB5
RT   (positions 552-788).";
RL   HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:841-850(1979).
RN   [5]
RP   SEQUENCE OF 809-947.
RX   MEDLINE; 80026030.
RA   DEWES H., FIETZEK P.P., KUHN K.;
RT   "The covalent structure of calf skin type III collagen. V. The amino
RT   acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A
RT   (position 789-927).";
RL   HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:851-860(1979).
RN   [6]
RP   SEQUENCE OF 948-1049.
RX   MEDLINE; 80026031.
RA   ALLMANN H., FIETZEK P.P., GLANVILLE R.W., KUHN K.;
RT   "The covalent structure of calf skin type III collagen. VI. The amino
RT   acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT   1(III)CB9B (positions 928-1028).";
RL   HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 360:861-868(1979).
CC   -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC       ALONG WITH TYPE I COLLAGEN.
```

```
CC  -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC      LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC      ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC  -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC      UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR  PIR: A02862; CGB07S.
DR  PROSITE; PS01208; VWFC; PARTIAL.
KW  EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW  GLYCOPROTEIN; COLLAGEN.
FT  CHAIN          1   1049       COLLAGEN ALPHA 1(III) CHAIN.
FT  DOMAIN         1     14       NONHELICAL REGION (N-TERMINAL).
FT  DOMAIN        15   1040       TRIPLE-HELICAL REGION.
FT  DOMAIN      1041   1049       NONHELICAL REGION (C-TERMINAL).
FT  MOD_RES       95     95       HYDROXYLATION.
FT  MOD_RES      107    107       HYDROXYLATION.
FT  MOD_RES      119    119       HYDROXYLATION.
FT  MOD_RES      938    938       HYDROXYLATION.
FT  MOD_RES      950    950       HYDROXYLATION.
FT  CARBOHYD     107    107       CROSS-LINK SITE.
FT  CARBOHYD     950    950       CROSS-LINK SITE.
FT  DISULFID    1040   1040       INTERCHAIN.
FT  DISULFID    1041   1041       INTERCHAIN.
SQ  SEQUENCE   1049 AA;  93651 MW;  7B3FA824 CRC32;

alignment_scores:
        Quality:   348.00             Length:   1139
        Ratio:       0.740            Gaps:       73
Percent Similarity:  41.264     Percent Identity:  26.163

alignment_block:
US-09-030-606-110/rev x CA13_BOVIN   ..

Align seg 1/1  to: CA13_BOVIN  from: 1   to: 1049

3051 AAACAGGGTGTGAGCTGGTGGGGAAAGTTGGGGTAGGGGAAAGTTGG 3002
     |||::|||   :::::||||||     |||      :::||
   8 LysSerGlyValAlaGlyLyIleAlaGlyTyrProGlyProAlaGl   24

3001 GGGTAGGGGAAATTTGGGCAGTGCCTTCATCAGCCCAGTCCTAGAGAGA 2952
     |   |||                           |||::|||
  24 yProProGly.................ProProGlyPro.ProGly   33

2951 GTAGAGGGG.........AGTGAAGTGGGGGAACCAGGCTGGGCC 2914
     :::::::||          |||::||    |||     |-
  34 ThrSerGlyHisProGlyAlaProGlyAlaProGlyTyrGlnGlyProPr 50

2913 AAGAGAAGAGGGGTGGTTAGGGGAAGCCGTTCAGACCTGAAGCCCACCCT 2864
     | |   |||    |||    ||||:::||::|||||||
  50 oGlyGluProGlyGlnAlaGly...ProAlaGlyProProGlyProProG 66

2863 CTACCTTCCTTCAACACCCTAACCTTGGGTAACACAGCATTTGGAATTATCA 2814
     ::|                                         :::
  66 ly.......................................Ala 67

2813 TTTGGGATGAGT.........AGAATTTCCAAGGTCCT 2785
     ::|:||  |||          ||| |||
  68 IleGlyProSerGlyLysAspGlyGluSerGlyArgProGly.ArgProG 84

2784 GGGTTAGGCATTTGGGGGGCCAGACCCCAGGAGAAGAAGATTCTGGCAA 2735
     |||  ||||||                    :::::||:::||||
  84 lyProArgGlyPheProGlyPro............AsnGlyAlaLysGlyAspAsp 91

2734 TGATCAGCCCAATGACCAGCTATCCCAGGGACCTGATTGTTGGGGATCC 2685
     :::::::|||              ||::::|||
  92 .........ProGlyMetLysGlyPro......AlaGlyMet 100

2684 CCCACCCTACCCAAATATTAGCACCAACAACAGA.........2651
     :::  |||::::|:||                    ||||:
 100 tProGlyPheProGlyMetLys...GlyHisArgGlyPheAspGlyArgA 116

2650 .........AAAGCTAGCAATGGATTCCCTCTACTTTGTTAAATAAATAA 2609
     |||:::::   |||  |||:::||
 320 snGlyLeuProGlyAlaAlaGlyGlyValProGlyPheArgGly 336
```

```
 116 snGlyGluLysGlyGlyGluProGlyAlaProGly..............126

2608 GTTAAATATTAAATGCCTGTGTCTCTGTGATGGCAACAGAGAGGACCAAC 2559
     :::|||    |||:||:::|||||:::|:|||::::|     :::::
 127 LeuLysGlyGluAsnGlyValProGlyGluAspGlyAlaProGlyProMe 143

2558 AGGCCACATCCTGATAAAAGTAAGAGGGGGGTGGATCAGCAAAAGACA 2509
     |||                      :::::||||:::||| :
 143 tGlyProArgGlyAlaProGlyGlyArgGlyArgProGlyLeuProGlyA 160

2508 GTGCTGTGGGCTGAGGGACCTGGTTCTTGTGTTGCCCTCAGGACTC 2459
     ::|||  |||:::  |||:::
 160 laAlaGlyAlaArgGlyAsn...........................166

2458 TTCCCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGTGTTTCAT 2409
                                                  166

2408 CCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGG 2359
                                                    |
 167 ..........................................AspG 168

2358 GGCTTAGA......GATGGGAAACCAGGTGACTGAGTTTATTCAGCTCCC 2315
     || |||       |||||||:::|||||||                |||
 168 lyAlaArgGlySerAspGlyGlnProGly...................Pro 178

2314 AAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCT 2265
     :::|||     |||:||||     :::
 179 ProGlyProPro..GlyThrAlaGlyPhePro..................188

2264 GAGCCTGGTAATCCACCTGCAGAGTCCCCAGCCATTCCAGTGCCATGGAGCC 2215
     |||:::|||||||||:::||::    ::: ||| ||||
 189 .......GlySerProGlyAlaLysGlyGluValGlyProAlaGlySerP 203

2214 CTTCTGGCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTC 2165
     ||:::
 203 roGly.............................................204

2164 CAGTCAGGCAGCCCTAGAGACTGGGGAGAG...........AGGAG 2130
     :::||||||:::||| |||||||
 205 SerSerGlyAlaProGlyGlnArgGlyGlyProGlyProGlnGlyHisAl 221

2129 AGGGACCGCCCAGCCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGG 2080
     ||||||:::::|| |||||||                   :::
 221 aGlyAlaProGlyProPro.....................GlyProProG 231

2079 GTGGCAGCAGAGCCACCATTACTTTGGCAGCAACAGAAACTGGCGGCCA 2030
     ||:::::::  :::||          |||:::::: :::|||
 231 lySerAspGlySerPro........GlyGlyLysGlyGluMetGlyPro 244

2029 GCCCGGCAGCCCATGGGGCTAACAGGAGCGGGGAGCTGGGACCCAGTGA 1980
     ||| ||  :::    :::                 |||   |||||
 245 AlaGlyIleProGlyAlaProGlyLeuIleGlyAlaArgGlyPro.....259

1979 GGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTTCACGCTGAGTATTTG 1930
     ::|||||                   ||:||||||||||     :::
 260 .ProGlyProProGlyThrAsnGlyValProGlyGlnArgGlyAlaAlaG 276

1929 GCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAGTAAATGGCGAC 1880
     |||  |||:::                :::   :::::|||:||:::|||
 276 lyGluProGlyLys...............AsnGlyAlaLysGlyAlaGlyAsp 286

1879 CAGACCCCAGGCCCTGCGGCAGCACCATATAGGCAGTGACAGACTGGCTGA 1830
     |||    :::   :::   |||||      ::::||
 287 ProGlyProArgGlyGluArgGlyLysAspGlySerProGlyIleAlaGl 303

1829 GCTGGCAAATGGAGCCCATAAACAGGGATG....................1800
     ::: |||                         |||::::||
 303 yProLysGlyGluArgGlyLysAspGlyLysSerProGlyValProGlyAlaA 320

1799 ..GGGCCACCTGGACAGCAGGA...AGGCACTACCAGGATGGCGAGGT 1755
     :::                       ||||:|:|||:||:::||||||||
 320 snGlyLeuProGlyAlaAlaGlyGlyValProGlyPheArgGly 336
```

992 .........AGGAGGGGGCGACAGCCCTTCTGCTGGCTGGTGGGG 955
600 GlyLysAspGlyProArgGlyProThr.................Gl 609

954 CCCAGCGCTGCCTCCTCAGCCACCAGCAGTGTGGCTGCTACCAGGTGAG 905
609 yProIleGlyProGlyProAla.......................G 618

904 GAAGATGAGGGTGAGCAGCAGCCAAAGAGGCACTCCTCCGGTGCCAGGT 855
618 lyGlnProGlyAspLysGlyGluSerGlyAla...ProGlyValProGly 633

854 ..AGGGGGCCAGGCACTGGTGTCCCAGTCAATGGCAGGCAGGAGGTAG 808
634 IleAlaGlyPro...............................ArgGlyGl 640

807 CCCAGGCAGCCCCCAAGACTGATCATGAAGGCATAGACAGAGGCCTG 758
640 yProGlyGlu..........................ArgGlyGlnGlyProProG 651

757 GCCACAGTGGT...CCGGGT...CCCGGAAGAGGTCAGAGAGCAGGGCCT 714
651 lyProAlaGlyPheProGlyAlaProGlyGlnAsnGlyGluProGlyAla 667

713 CCAGTGGAGTGAAGCACACCTGCCACAGAAGTCCAGCAGCCCCACGCCC 664
668 LysGlyGluArgGlyAlaProGlyGluLysGlyGluGlyGlyPro...Pr 683

663 AGGATGAGCAGTGGCCAGCTCCAGGG............GCCTGGGATCCGGGCA 623
683 oGlyAlaAlaGlyProAlaGlySerGlyProAlaGlyProProGlyP 700

622 CAGCAGCCCTGCTAGCCAGCCGCGCCCTTGGGATGAGAAAGAGGCTCAGCA 573
700 roGlnGlyValLysGly.............GluArgGlySerPro 710

572 GGATGCCCAAGGACAGTGCCCAGATGAGGGCCGGC........GGCGGCCA 529
711 GlyGlyProGlyAla........................AlaGlyPheProGlyGlyAr 722

528 TAGCGTCCACGCCAGTGGTCACTGGCGCTGCAGGAGGCGGGACACAGAC 479
722 ....ProGlyProProGlySerAsn.........ProP 725

478 CAGGCCCAGCACTGGACCAATGCCAGCACCAATGGTCATGAACTTCCT 429
725 roGlyPro..................................ProGlySerAsnGlyAsnPro 734

428 CTACCCCA........................CTTCCAGCAGCAGAGGCCGCACATAGGTG 391
735 GlyProProGlySerSerGlyAlaProGlyLysAspGly........... 747

390 ATGCCTGCGCCAAACACCTCCAGGCCCAAAGGTTAGCAGGTTGACCAG 341
748 ....ProProGlyProProGlySerAsn.........GlyAlaProG 759

340 CAAGAGCTGGCTTCCGGTGCCGCAGCAGGC............................. 309
759 ....lySerProGlyIleSerGlyProGlyAspSerGlyProGlyGlu 775

308 ..GGCTCACCCACAGCCCTGACCCATAGTGGG...CCAGGCGGGTAGG 265
776 ArgGlyAlaProGlyProGlnGlyPro.ProGlyAlaProGlyProLeuG 792

264 GCTCAGGGGGCGTTCAGGCACTCCAGAACTGCTTCGTCCGGCGCTCTGCT 215
792 lyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProProGlyMet 808

214 CCAGAAAGCTGCGGCCTCTCCT............................. 194
809 ProGlyAlaArgGlySerProGlyProGlnGlyIleLysGlyGluAsnGl 825

193 .........CCTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCC 154

```
825  yLysProGlyProSerGlyGlnAsnGlyEuArgGlyProProGly....  840
          |||                    ::::::       |||||

153  ATTTCTGCCAGCCCTTGGT.............GCCGGTCCAGCTTCTCAG  116
          |||                ||||||      ||:::::||
841  ........ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGlyGlu  853

115  CCCATGCTC.....AACACCTGCTGCTGTGGGCACCCAGTGGGACA  72
          |||      ||| ||::::::|||      |||   |||
854  ProGlyArgAspGlyAsn.ProGlySerAspGlyLeuProGlyArgAspG  870

71   CGTCTCATCCACTCAGATCCTGGCCAGGCCGCGGCGCTGTCACCCGGAGCC  22
          :::          ||||||       :::  |||||||
870  lyAlaProGlyAlaLysGlyAspArgGlyGluAsnGlySerProGlyAla  886

seq_name: SwissProt_37:CA11_CHICK

seq_documentation_block:
ID  CA11_CHICK    STANDARD;      PRT;   1453 AA.
AC  P02457;
DT  21-JUL-1986 (REL. 01, CREATED)
DT  01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT  01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE  PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN  COL1A1.
OS  GALLUS GALLUS (CHICKEN).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC  NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN  [1]
RP  SEQUENCE OF 1-153 FROM N.A.
RX  MEDLINE; 88055316.
RA  FINER M.H., BOEDTKER H., DOTY P.;
RT  "Construction and characterization of cDNA clones encoding the 5' end
RT  of the chicken pro alpha 1(I) collagen mRNA.";
RL  GENE 56:71-78(1987).
RN  [2]
RP  SEQUENCE OF 1-144 FROM N.A.
RX  MEDLINE; 88007542.
RA  FINER M.H., AHO S., GERSTENFELD L.C., BOEDTKER H., DOTY P.;
RT  "Unusual DNA sequences located within the promoter region and the
RT  first intron of the chicken pro-alpha 1(I) collagen gene.";
RL  J. BIOL. CHEM. 262:13323-13332(1987).
RN  [3]
RP  SEQUENCE OF 152-1187.
RX  MEDLINE; 82231995.
RA  HIGHBERGER J.H., CORBETT C., DIXIT S.N., YU W., SEYER J.M.,
RA  KANG A.H., GROSS J.;
RT  "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the
RT  complete primary structure of the helical portion of the chick skin
RT  collagen alpha 1(I) chain.";
RL  BIOCHEMISTRY 21:2048-2055(1982).
RN  [4]
RP  SEQUENCE OF 1200-1205.
RX  MEDLINE; 72243016.
RA  EYRE D.R., GLIMCHER M.J.;
RT  "Evidence for a previously undetected sequence at the carboxyterminus
RT  of the alpha 1 chain of chicken bone collagen.";
RL  BIOCHEM. BIOPHYS. RES. COMMUN. 48:720-726(1972).
RN  [5]
RP  SEQUENCE OF 981-1453 FROM N.A.
RX  MEDLINE; 81160715.
RA  FULLER F., BOEDTKER H.;
RT  "Sequence determination and analysis of the 3' region of chicken pro-
RT  alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT  including the carboxy-terminal propeptide sequences.";
RL  BIOCHEMISTRY 20:996-1006(1981).
RN  [6]
RP  SEQUENCE OF 1311-1453 FROM N.A.
RX  MEDLINE; 80134546.
RA  SHOWALTER A.M., PESCIOTTA D.M., EIKENBERRY E.F., YAMAMOTO T.,
RA  PASTAN I., DECROMBRUGGHE B., FIETZEK P.P., OLSEN B.R.;
RT  "Nucleotide sequence of a collagen cDNA fragment coding for the
RT  carboxyl end of pro alpha 1(I)-chains.";
```

```
RL  FEBS LETT. 111:61-65(1980).
CC  -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC      (FIBRILLAR FORMING COLLAGEN).
CC  -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC  -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC      BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC      HYDROXYAPATITE.
CC  -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC      UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC  -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute.  There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed.  Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; M17839; G555435; -.
DR  EMBL; M17838; G555435; JOINED.
DR  EMBL; V00401; G63308; -.
DR  EMBL; M10571; G555443; ALT_SEQ.
DR  EMBL; M17607; G211474; -.
DR  PIR; A02857; CGCH1S.
DR  PIR; A27179; A27179.
DR  PIR; A29367; A29367.
DR  PROSITE; PS01208; VWFC; 1.
DR  PFAM; PF00093; vwc; 1.
KW  EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW  GLYCOPROTEIN; COLLAGEN; SIGNAL.
FT  SIGNAL      1     22
FT  PROPEP     23    151              AMINO-TERMINAL PROPEPTIDE.
FT  CHAIN     152   1205              COLLAGEN ALPHA 1(I).
FT  PROPEP   1206   1453              C-TERMINAL PROPEPTIDE.
FT  DOMAIN     31     89              VWFC.
FT  MOD_RES   152    152              PYRROLIDONE CARBOXYLIC ACID.
FT  MOD_RES   254    254              HYDROXYLATION (POTENTIAL).
FT  MOD_RES   851    851              HYDROXYLATION (POTENTIAL).
FT  MOD_RES  1081   1081              HYDROXYLATION (POTENTIAL).
FT  MOD_RES  1097   1097              HYDROXYLATION (POTENTIAL).
FT  MOD_RES  1153   1153              HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT                                    ONLY HYDROXYLATED PRO IN POSITION X (IN
FT                                    THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
FT  CONFLICT  1187   1187            F -> L (IN REF. 5).
FT  CONFLICT  1441   1441            Q -> H (IN REF. 6).
SQ  SEQUENCE  1453 AA;   137789 MW;  5DAC8ADE CRC32;

alignment_scores:
     Quality:           346.00             Length:          1049
     Ratio:               0.810             Gaps:              64
Percent Similarity:     40.705       Percent Identity:    26.787

alignment_block:
US-09-030-606-110/rev x CA11_CHICK

Align seg 1/1  to: CA11_CHICK  from: 1  to: 1453

2551  ATCCTGATAAAAGGTAAGAGGGGGTGGATCAGCAAAAGACAGTGCTGT  2502
          ::|||    ::|||:::        |||   ||| |||||||:||
18    LeuLeuThrArgGlyGluGlyGluGlnThrGlySerCysVa  34

2501  GGGCTGAGGG............................GACCTGG  2485
          :::||||                            :::
34    lGlnAspGlyLeuThrTyrAsnAspLysAspVaLTrpLysProGluProc  51

2484  TTCTTGTGTGT...TGCCCCTCAGGACTCTTCCCTACAAATAAGTCATA  2438
          ::|||||  ||| |||||||      :::||
51    ysGlnIleCysValCysAspSerGlyAsnIleLeuCysAspGluValIle  67

2437  TGTTCAAAT...............................CCCATGGAGGAGTG  2415
          ||:::::                              |||||
```

```
  68 CysGluAspThrSerAspCysProAsnAlaGluIleProPheGlyGluCy  84

2414 TTTCATCTA.....GAAACTCCCATGCAAGAAGCTACATTAAACGAAGC  2371
              ::: ||||:::
  84 sCysProIleCysProAspValAspAlaSerProValTyrProGluSerA  101

2370 TGCAGGTTAAGGGGCTT......                             2354
     |||:::|||
 101 laGlyValGluGlyProLysGlyAspThrGlyProArgGlyAspArgGly  117

2353 ..........AGAGATGGGAAACCAGGTGACTGAGTTTATTC          2322
              |||||||| ||||||
 118 LeuProGlyProGlyArgAspGlyIleProGlyGlnProGlyLeuPr     134

2321 AGCTCCAAAAACCCTTCTCAGGTGTGTCTCAACTAGGAGGCTAGCTGT    2272
     :::||| |||||||
 134 oGlyProProGlyProPro...GlyProProGlyLeuGlyLeuGlyAsnPheA 150

2271 TA...ACCCTGAGCCTGGGT.........AATCCACCTGCA           2243
        ::::|| |||||
 150 laProGlnMetSerTyrGlyTyrAspGluLysSerAlaGlyValAlaVal  166

2242 GAGTCCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGT  2193
     :::||| |||  |||||||||
 167 ProGlyProMetGlyProAla..GlyProArgGlyLeuPro.........  179

2192 CCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGCAGCCCTAGAGACT  2143
                             ||| |||
 180 ..............GlyProPro......                       182

2142 GGGGAGGAGGAGGAGGACGCCCCCAGCCGTGTCCAGCTACGCCACC     2093
               ::||||::::
 183 ..............GlyAlaProGlyProGlnGlyPheGln....       191

2092 TCAGCAGCACAGGTGGCAGCAGAGCCACATTACTTGGCAGCAACAG     2043
     :::
 192 ...GlyProProGlyGluProGlyProAlaSerGlyProMetGl        207

2042 AAACTGGCGCGCAGCCCGCAGCCCATGGGGCTAACAGGAGGCGGGGAGC  1993
     ||||||| :::|||||
 207 yProArgGlyProAlaGlyProProGlyLysAsnGlyAspAspGlyGluA  224

1992 TGGGA.....CCCAGTGAGGCAGGCCCTCCA..........           1967
     |||         ||||||||
 224 laGlyLysProGlyArgProGlnArgGlyProProGlyProGlyGlnGly  240

1966 ..........CCCCAATGTGCTGGAAG..........              1950
               ||| |||||
 241 AlaArgGlyLeuProGlyThrAlaGlyLeuProGlyMetLysGlyHisAr  257

1949 ....TTTTCTACGTGAGTATTTGGCCAAGTCGCTCTTGTCAAATACTAC  1904
         ||| |||||
 257 gGlyPheSerGlyLeuAsp......                           263

1903 CTGTGTAGCAAAGTAAATGGCGACCAGACCCA...GGCCTGCGGCAGACA 1857
     ::||| :::|| ||
 264 ......GlyAlaLysGlyGlnProGlyProAlaGlyProGlyLysGlyGlu 277

1856 CCATATAGGCAGTGACAGACTGGCTGGACAATGGAGCCCATAAAC      1807
     ::::|||| |||||| ::||||
 278 ProGlySerProGlyGluAsnGlyAlaProGlyGlnMetGlyPro...... 292

1806 AGGGGATGGGGCCACCTGGGACAGCAGGAGGCACTATCCA......     1767
     ||| |||||| |||||
 293 ...ArgGlyLeuProGlyGluArgGlyArgProGlyProSerGlyProA   308

1766 ..GGATGGCGAGGTCACAGCCAGGATGCCCCGGCCCGGAA...CCACCCTGG 1722
     ||||||||||| :::|||||
 308 laGlyAlaArgGlyAsnAspGlyAlaProGlyAlaAlaGlyProProGly  324

1721 CCTCCGGTGGGCTCACCCACCCACCA............CG           1694
     ||||||||
 325 ProThrGlyProAlaGlyProProGlyPheProGlyAlaAlaAlaGlyAlaLy 341
```

```
1693 TACGGAGACATCACAGGCAGCAGAGGCCCCAGAGAGCCGGGTGGAGGTGGGA 1644
     ::: ||||||:::
 341 sGlyGluThrGlyProGlnGlyAlaArgGlySerGluGlyProGlnGlyLys 358

1643 GCAGG.....CCACTGCCTCCAGCACCACCCAGTGTCCATTAGGG...AAG 1603
     ||||      ||| |||||:::||:::||
 358 erArgGlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAla  374

1602 GGAGCTCCAGGCTTAGGG.....CCTGGCAGGAAGCTGGTCATCAGGCT   1559
     ||||||| |||:::||
 375 GlyAsnProGlyAlaAlaAspGlyGlnProGlyAlaLys.GlyAlaThrGlyA 391

1558 GTCCTCACTGCTAGCACCTCCAGTGTCCCCTCGTATTTGGCCAGGAACA  1509
     ::||| |||||.||||||
 391 laPro......GlyIle..AlaGlyAla                        397

1508 CCTGCTTCTCCCGGTGGTAGAGGG.....AGGCCAGTGTGTAGGGCAGG  1465
     |||::: |||||:||||| :::||
 398 ProGlyPheProGlyAlaArgGlyPheSerGlyProGlnGlyProSerGl  414

1464 ATCTGCAGGGTGAGAAGTGAACCCGGTGAACCCGTGAGCCTGTCAC     1415
     :::||| ||  ::  ||||| ||||| ||
 414 yAlaProGlyProLys.GlyAsnSerGlyGluProGlyAlaProGlyAsn  430

1414 CACGGCCACACT............                            1403
     ||| |||
 431 LysGlyAspThrGlyAlaLysGlyGluProGlyProAlaGlyValGlnGlnGlI 447

1402 ..........GTGGGACAGGCATGTGGCACCGGCAGCCACAGGGAAA     1366
               ::||||||||:::
 447 yProProGlyProAlaGlyGluGlyUGlyLysGlyAlaArgGlyAlaArgGlyGlUp 464

1365 GCTGCCACACTGGCCAAATAGACTGCTGCGAGTGCCGA.....        1329
     :::
 464 roGlyProAlaGlyLeuProGlyProAlaGlyGluArgGlyAlaProGly  480

1328 ...........ATCGCTGCACCACCGGTCCATG                   1306
                ||||||||||||| |||||
 481 SerArgGlyPheProGlyAlaAspGlyIleAlaGlyProLysGlyProPr  497

1305 ACCAGAGAGA............                              1296
     ||||||
 497 oGlyGluArgGlySerProGlyAlaValGlyProLysGlySerProGlyG  514

1295 ..........AGACCAGGAGATGGCGCACTGCAGGAACA.            1266
               ::::
 514 luAlaGlyArgProGlyGluAlaGlyProGlyAlaLysGlyLeuThr     530

1265 ...GCCCCAGGCTGCCATCCGAACGCCTTCATCATAGTGTCTCCGGGGC  1219
     ::: |||||||| ||||
 531 GlySerProGlySerProGlyPro........AspGlyLysThrGlyPr   544

1218 TCGGTGCCCGGCTCAGCTCTGGGCACGCCCTGTACAGCCCCTCGCCCAC  1169
     ||||| ::||||||||| |||
 544 oProGlyProAlaGlyIleGlyAspGlyArgProGlyProAlaGlyProPro. 560

1168 GAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCAGCTGCACA 1119
     |||
 560 ......................                             560

1118 GCTCAGCCACGAAGAGCCGGCACGGCCGCAGGGTGCGGGCATGCGGCAGCAGAGC 1069
     ||||:: ||||| :::
 561 ......GlyAlaArgGlyGlnAlaGlyValMetGlyPhePheProGlyProLy 575

1068 TGGTGCAGCCGGGGAAGCAGGGCCGCCCAGGTTCC...GGAAAGCCAAGCG 1022
     |||||||||||| :::|||||||
 575 sGlyAlaAlaGlyGlyLuProGlyLysProGlyLeuAlaProGlyProGlyP 592

1021 GGCCCGGCCATGGACAGCAGTGGGGCGACAAGGAGGGGGCCGCCAGCCCTT 972
     ||||
 592 roProGly......AlaValGlyAlaAlaAlaGlyLysLysAspProGlyGluAla. 605
```

```
971 CTGCTGGCTCGGTGGGCCCCAGCGCGCTGCCTCCTCAGCCACCAGCAGTGTG 922
            |||             |||                |||
606 ...............GlyAlaGlnGlyProGlyProThr.......... 614

921 GCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCACT.. 873
            |||   ||||||::||||
615 ..............GlyProAlaGlyGluArgGlyGluG 623

872 .....CCTCCTGGGTGCCCAGGT..AGGGGG...CCAGGGCACTGGTGT 834
            |||  |||:::|||       :::          |||||
623 lnGlyProAlaGlyAlaProGlyPheGlnGlyLeuProGlyProAlaGly 639

833 CCCAGTCAATGGCAGGCAGGAGGTAGCCCCAGGCAGCCCCCAGAGACTGATC 784
            |||   |||::|||||||         :::||
640 ProPro...........GlyGluAlaGlyLysProGlyGlu...... 649

783 ATGAAGGCATAGACAGAGTAGGCCTGGCACAGTGGT...CCGGGTCCC. 738
            |||::||::|||::|||   |||||||||   ||||||||||
650 ............GlnGlyValProGlyAsnAlaGlyValAlaProGlyProA 662

737 ...GGAAGAGGTCAGAGAGCAGGAGCCCTCCAGTGGAGTGAAGCCACCT 693
            ::||||||||::||   ||   ||| |:   ::||||
662 laGlyAlaArgGlyGluArg...GlyPheProGlyGluArgGlyValGln 677

692 GGCCACGAGAAGTCCAGCAGCCCCACCCCAGGATGAGCAGTGCCA...... 648
            |||   ::::||||   ::|||||   ||||||   ::: |||
678 GlyProProGlyProGlnGlyProArgGlyAlaAsnGlyAlaProGlyAs 694

647 ..........GCTCCAGGGGCCTGGGATCCGGGCA 623
                      ||||||||||||       |||||||
694 nAspGlyAlaLysGlyAlaAspAlaGlyAlaProGlyAla......ProGlyA 709

622 CAGCAGCCCCTGCTAGCCAGCCGCCCCTTGGGATGAGAAAGAGGCCTCAGCA 573
            ::   ::   :|||:::||   ||||||||   ||||||||||||||
709 snGluGlyProGlyLeuGluGlyMetProGlyGluArgGlyValAlaAla 725

572 GGATCCCA.............AGGACAGTGCCCAGGATGAAGGGCCGG 538
            |||  |||               :::::||
726 GlyLeuProGlyAlaLysGlyAlaAspArgGlyAlaAspGlyProLysGlyAl 742

537 CGGCGGGC.............. 531
            ||                    ||
742 aAspGlyAlaProGlyLysAspGlyAspGlyLeuArgGlyLeuThrGlyProIleG 759

530 .........CATAGCGTCCACCCCAGTGTCACTGGCTGAGCTGAGGACGG 489
            ::|||   ||||||   ||||||::::||
759 lyProProGlyProAlaGlyProGly.........AspLysGlyGlu 772

488 GGACACAGCAGGCCCAGCA.........CTGGACCAATGCCCAGCACCATG 445
            :::   ||||||||:||||                        ||||||
773 AlaGlyProProGlyProAlaGlyProThrGlyAlaArgGlyAlaProGly1 789

444 GTCATGAACTTCTCCTCCTACCCCCACTTCCAGCAGCAGAGGCGGCACATA 395
            |||::||| |||   |||||::: ||||||   |||::
789 yAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly.... 804

394 GGTGATGCCTGCGGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGA 345
            ||||| |||:|||||              |||||||
805 ........ProProGly......AlaAspProGlyGln 811

344 CCAGCAAGAGCTGGGCTTTCCGGT...GCCGCAGCAGGCGGCTCACCCAC 298
            |||:::||| ||| ||:::||   |||::: |||||
812 ProGlyAlaLysGlyGluThrGlyAspAlaLysGlyAlaLysGlyAspAlaGl 828

297 AGCCTCTGGACCATAGTGGGCCA...GGCGGGTAGGCCTCAGGGGGCCGT 251
            ||||| :::|||||   ||||||    ||| ||:::
828 yProProGlyPro.AlaGlyProThrGlyAlaArgGlyAlaProAlaGly*** 844

250 TCAGGCCACTCCAGAACTGCTTCGTCTCGGCTCCTCGGCTCCTCCAGGAATCAG 203
            ::|||:::|||     :::||
845 ValGlyAlaPro.............GlyProLysGlyAlaAr 854

202 ......GCCTCTCCCTCCTTGCTCGCCGCCAACTGCCTAGGAATCAG.... 164
```

```
854 gGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyA 871
            |||:::|||||||      :::|||
163 ..............CCAGGCGCCCATTCTGCCAGCCCCTTGGTGCC 131

871 rgValGlyProProGlyProSerGlyAsnIleGlyLeuProGlyProPro 887
            |||:::::::::||||          |||
130 GGTCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGGGCACCTCA 81

888 GlyProAlaGlyLys***GlySerLys.............GlyProAr 899
            |||||||::||||           ::              :::||||
80 GTGGGGACACGTCTCATCACTCAGATCCTGGGCCGA........... 46

899 g.GlyGluThrGlyProAlaGlyArgProGlyGluProGlyProAlaGly 915
            : |||||||||||        ::: |||||||||
45 ..........GGCGGCGGCGGCTGTCACCCGGAGCC 22

916 ProProGlyProGlyGluLysGlySerProGlyAla 928
            ||    :::|||||||||||
22
```

seq_name: SwissProt_37:CA11_HUMAN

```
seq_documentation_block:
ID   CA11_HUMAN     STANDARD;      PRT;   1464 AA.
AC   P02452; Q15176;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.
GN   COL1A1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE OF 1-472 FROM N.A.
RX   MEDLINE; 89025644.
RA   TROMP G., KUIVANIEMI H., STACEY A., SHIKATA H., BALDWIN C.T.,
RA   JAENISCH R., PROCKUP D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT   chain of human type I procollagen.";
RL   BIOCHEM. J. 253:919-922(1988).
RN   [2]
RP   SEQUENCE OF 1-181 FROM N.A.
RX   MEDLINE; 84270697.
RA   CHU M.-L., DE WET W.J., BERNARD M.P., DING J.-F., MORABITO M.,
RA   MYERS J., WILLIAMS C., RAMIREZ F.;
RT   "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT   conservation of a pattern of introns and exons.";
RL   NATURE 310:337-340(1984).
RN   [3]
RP   SEQUENCE OF 162-301.
RX   MEDLINE; 71038625.
RC   TISSUE=SKIN;
RA   CLICK E.M., BORNSTEIN P.;
RT   "Isolation and characterization of the cyanogen bromide peptides from
RT   the alpha 1 and alpha 2 chains of human skin collagen.";
RL   BIOCHEMISTRY 9:4699-4706(1970).
RN   [4]
RP   SEQUENCE OF 263-268.
RC   TISSUE=SKIN;
RX   MEDLINE; 71001508.
RA   MORGAN P.H., JACOBS H.G., SEGREST J.P., CUNNINGHAM L.W.;
RT   "A comparative study of glycopeptides derived from selected
RT   vertebrate collagens. A possible role of the carbohydrate in fibril
RT   formation.";
RL   J. BIOL. CHEM. 245:5042-5048(1970).
RN   [5]
RP   SEQUENCE OF 425-1464 FROM N.A.
RX   MEDLINE; 84080385.
RA   BERNARD M.P., CHU M.-L., MYERS J.C., RAMIREZ F., EIKENBERRY E.F.,
RA   PROCKOP D.J.;
RT   "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT   pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT   of structures that are conserved during evolution.";
```

RL   BIOCHEMISTRY 22:5213-5223(1983).
RN   [6]
RP   SEQUENCE OF 1229-1454 FROM N.A.
RC   TISSUE=BONE;
RX   MEDLINE; 88124208.
RA   MAEKELAE J.K., RAASSINA M., VIRTA A., VUORIO E.;
RT   "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT   domain.";
RL   NUCLEIC ACIDS RES. 16:349-349(1988).
RN   [7]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE; 88097389.
RA   BORNSTEIN P., MCKAY J., MORISHIMA J.K., DEVARAYALU S., GELINAS R.E.;
RT   "Regulatory elements in the first intron contribute to
RT   transcriptional control of the human alpha 1(I) collagen gene.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 84:8869-8873(1987).
RN   [8]
RP   SEQUENCE OF 1-34 FROM N.A.
RX   MEDLINE; 85130970.
RA   CHU M.-L., DE WET W.J., BERNARD M.P., RAMIREZ F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. BIOL. CHEM. 260:2315-2320(1985).
RN   [9]
RP   SEQUENCE OF 1-44 FROM N.A.
RX   MEDLINE; 88033098.
RA   ROSSOUW C.M.S., VERGEER W.P., DU PLOOY S.J., BERNARD M.P., RAMIREZ F.,
RA   DE WET W.J.;
RT   "DNA sequences in the first intron of the human pro-alpha 1(I)
RT   collagen gene enhance transcription.";
RL   J. BIOL. CHEM. 262:15151-15157(1987).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 91184577.
RA   KUIVANIEMI H., TROMP G., PROCKOP D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 97255959.
RA   KUIVANIEMI H., TROMP G., PROCKOP D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   HUM. MUTAT. 9:300-315(1997).
RN   [12]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE; 91374476.
RA   BYERS P.H., WALLIS G.A., WILLING M.C.;
RT   "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL   J. MED. GENET. 28:433-442(1991).
RN   [13]
RP   REVIEW ON OI VARIANTS.
RX   MEDLINE; 97169389.
RA   DALGLEISH R.;
RT   "The human type I collagen mutation database.";
RL   NUCLEIC ACIDS RES. 25:181-187(1997).
RN   [14]
RP   VARIANT OI-II CYS-1166.
RX   MEDLINE; 86287390.
RA   COHN D.H., BYERS P.H., STEINMANN B., GELINAS R.E.;
RT   "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT   change in one human pro alpha 1(I) collagen allele.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 83:6045-6047(1986).
RN   [15]
RP   VARIANT OI-II ARG-569.
RX   MEDLINE; 87222295.
RA   BATEMAN J.F., CHAN D., WALKERS I.D., ROGERS J.G., COLE W.G.;
RT   "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT   arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT   collagen.";
RL   J. BIOL. CHEM. 262:7021-7027(1987).

RN   [16]
RP   VARIANT OI-II CYS-926.
RX   MEDLINE; 88033031.
RA   VOGEL B.E., MINOR R.R., FREUND M., PROCKOP D.J.;
RT   "A point mutation in a type I procollagen gene converts glycine 748
RT   of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT   a lethal variant of osteogenesis imperfecta.";
RL   J. BIOL. CHEM. 262:14737-14744(1987).
RN   [17]
RP   VARIANT OI-II ARG-842.
RX   MEDLINE; 88299828.
RA   BATEMAN J.F., LAMANDE S.R., DAHL H.H., CHAN D., COLE W.G.;
RT   "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT   chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT   the peptide defect by in vitro expression of the mutant cDNA.";
RL   J. BIOL. CHEM. 263:11627-11630(1988).
RN   [18]
RP   VARIANT OI CYS-1195.
RX   MEDLINE; 89218628.
RA   LABHARD M.E., WIRTZ M.K., POPE F.M., NICHOLS A.C., HOLLISTER D.W.;
RT   "A cysteine for glycine substitution at position 1017 in an alpha
RT   1(I) chain of type I collagen in a patient with mild dominantly
RT   inherited osteogenesis imperfecta.;
RL   MOL. BIOL. MED. 5:197-207(1988).
RN   [19]
RP   VARIANT OI-II VAL-434.
RX   MEDLINE; 89255493.
RA   PATTERSON E., SMILEY E., BONADIO J.;
RT   "RNA sequence analysis of a perinatal lethal osteogenesis imperfecta
RT   mutation.";
RL   J. BIOL. CHEM. 264:10083-10087(1989).
RN   [20]
RP   VARIANT OI-IV SER-1010.
RX   MEDLINE; 89308591.
RA   MARINI J.C., GRANGE D.K., GOTTESMAN G.S., LEWIS M.B., KOEPLIN D.A.;
RT   "Osteogenesis imperfecta type IV. Detection of a point mutation in
RT   one alpha 1(I) collagen allele (COL1A1) by RNA/RNA hybrid analysis.";
RL   J. BIOL. CHEM. 264:11893-11900(1989).
RN   [21]
RP   VARIANTS OI-II ALA-1106; VAL-1151; ARG-1154 AND VAL-1184.
RX   MEDLINE; 89380165.
RA   LAMANDE S.R., DAHL H.-H.M., COLE W.G., BATEMAN J.F.;
RT   "Characterization of point mutations in the collagen COL1A1 and
RT   COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";
RL   J. BIOL. CHEM. 264:15809-15812(1989).
RN   [22]
RP   VARIANT OI SER-1022.
RX   MEDLINE; 90062068.
RA   PACK M., CONSTANTINOU C.D., KALIA K., NIELSEN K.B., PROCKOP D.J.;
RT   "Substitution of serine for alpha 1(I)-glycine 844 in a severe
RT   variant of osteogenesis imperfecta minimally destabilizes the triple
RT   helix of type I procollagen. The effects of glycine substitutions on
RT   thermal stability are either position of amino acid specific.";
RL   J. BIOL. CHEM. 264:19694-19699(1989).
RN   [23]
RP   VARIANT OI-II CYS-1082.
RX   MEDLINE; 89109573.
RA   CONSTANTINOU C.D., NIELSEN K.B., PROCKOP D.J.;
RT   "A lethal variant of osteogenesis imperfecta has a single base
RT   mutation that substitutes cysteine for glycine 904 of the alpha 1(I)
RT   chain of type I procollagen. The asymptomatic mother has an
RT   unidentified mutation producing an overmodified and unstable type I
RT   procollagen.";
RL   J. CLIN. INVEST. 83:574-584(1989).
RN   [24]
RP   VARIANT OI CYS-272; CYS-704 AND CYS-796.
RX   MEDLINE; 90009313.
RA   STARMAN B.J., EYRE D., CHARBONNEAU H., HARRYLOCK M., WEIS M.A.,
RA   WEISS L., GRAHAM J.M., BYERS P.H.;
RT   "Osteogenesis imperfecta. The position of substitution for glycine by
RT   cysteine in the triple helical domain of the pro alpha 1(I) chains of
RT   type I collagen determines the clinical phenotype.";
RL   J. CLIN. INVEST. 84:1206-1214(1989).

RN   [25]
RP   VARIANT OI-II CYS-422.
RA   FERTALA A., WESTERHAUSEN A., MORRIS G.M., ROONEY J.E., PROCKOP D.J.;

alignment_scores:
Quality: 343.00   Length: 952
Ratio: 0.858   Gaps: 62
Percent Similarity: 42.017   Percent Identity: 42.017

alignment_block:
US-09-030-606-110/rev x CA11_HUMAN

Align seg 1/1 to: CA11_HUMAN   from: 1 to: 1464

```
1695  ................................CGTACGGAGACATCACAG.................GCA  1675
 347  lyAlaValaGlyAlaLysGlyGluAlaGlyLysGlyGluAlaGlyLysProGlnGlyProrgGlySer  363

2353  AGAGATGGGAAACCAGTGACTGAGTTTATTCAGCTCCCAAAAACCCT..  2306
 131  ArgAspGlyIleProGlyLeuProGlyLeuProGlyProGlyProPr   147

1674  GAGGGCCCCAGAGCGCGGGTGGAGGTGGGAGCGCAGGCCACTGCCTCCAGC  1625
 364  GluGlyProGlnGlyValArgGly.........GluProGlyProProGl  377

2305  ...TCTCTAGGTGTGTCTCAACTAGGAGGCTAGCTGTTA...ACCCTGA  2263
 147  oGlyProProGlyProProGlyLeuGlyLysGlyAsnPheAlaProGlnLeuS  164

1624  ACCCACGTGTCCATTAGGG...AAGGAGCTCCAGGCTTAGG......C  1584
 377  yProAlaGlyAlaAlaGlyProAlaGlyAsnProGlyAlaAspGlyGlnP  394

2262  GCCTGGGTAATCCACCTGCAGAGTCCCCGCATTCCAGTGCATGGAGCCCT  2213
 164  erTyrGlyTyrAspGluLysSerThrGlyGlyIleSerValProGlyPro  180

1583  CTGGCAGGAGCTGGTCATCAGGCTGTCTCCACTGCTAGCACCTCCAGTG  1534
 394  yProAlaGlyAlaAlaGlyProGlnGlyAlaAsnGlyAlaPro          403

2212  TCTCGGCCTCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCA  2163
 180  Pr   183

1533  TCCCCTCGGTATTGGGCAGGAACACCTGCTTCTCCCGGTGTAGAGGGA  1484
 404  ......GlyIle..AlaGlyIleAlaProGlyPheProGlyAlaArgGly.  416

2162  GTCAGGCAGCCCTAGAGACTGGGAGAGAGGAGGAGGACGCCCCAGCCCC  2113
 183  oSerGlyPro.........................ArgGlyLeuProGlyProP  193

1483  GGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGA  1434
 416  .......ProSerGlyPro.GlnGlyProGlyGlyProGlyProGlyPr   429

2112  CAGCTGTGCAGCTACGCACCTCAGCAGCAGGGTGCAGCAGAGACCA......  2063
 193  roGlyAlaPro...GlyProGlyProGlyProProGlyProProGluPro   208

1433  GGGCGGCTGAAGCTGTCACCACGGCCACACTGGGACAGGCATGTGGCA  1384
 429  oLysGlyAsnSerGlyGluProGly..AlaProGlySerGlyGlyAspT   445

2062  CATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCCATGG  2013
 208  roGlyValGlnGlyPheGlnGlyProAlaGlyGluGluGlyLysArgGly  225

1383  CCGGCAGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT  1334
 445  hrGly.AlaLysGlyGluProGlyPro.........................Va  454

2012  GGCTAACAGGAGCGGGGAGCTGGGA.........CCCA  1984
 225  GlyGluProGlyLysSerGlyAspAspGlyGluAlaGlyLysLysProGlyArgProG  242

1333  GCCGAATCGCTCACCAGCCGTCCATGACCAGAG......AGAAGACCA  1290
 454  l.GlyValGlnGlyProGlyProAlaGlyGluGluGlyLysArgGly  470

1983  GTGAGGCAGGCCCTCCA.........CCCCAATGT  1958
 242  lyGluArgGlyProProGlnGlyAlaAlaArgGlyLeuProGlyThr  258

1289  GGGAGATGCGGCACTGCAGGAACAGCC..CCAGGCTGCCCATCCGAACG  1243
 470  AlaArgGlyAlaGlyProGlyProThrGlyLeuProGlyProGlyGluAr  487

1957  GCTCGAAGTTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTCAAATA  1908
 258                 ......CCTGGGACAGCAGGAAGCCACTATCCA  1767
 287  lyGluProGlySerProGly.........GluAsnGlyAlaProGly   299

1242  C...........................CTTCATCATAGTGTCTCCGGGC  1220
 487  gGlyGlyProGlySerArgGlyPheProGlyAlaAspGlyValAlaGlyP  504

1907  CTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGGCCTGCGGCAGAC  1858
 270  eSerGlyLeuAspGlyLysGlyAlaLysGlyAspAlaGlyProAlaGlyProLysG  287

1219  CTCGGTGCCGGCTCAGCTCTGGGCACCGCCCTGGTACAGCCCCTCGCCCA  1170
 504  roLysGlyProAlaGlyLeuArgGlySerProGlyProAlaGlyProLys  520

1857  ACCATATAGGACGAGTGACAGACGTGGCTGAGCTGACAATGAGCCCATAAA  1808
 287  ...........GluAsnGlyAlaProGly   299

1169  CGAAATCCGTGTAAAACGCGTGAAGGTCATGAGTGCCATCCAGCTGCAC  1120
 520  laGlyArgProGlyGluArgGlyGluSerProGlyProAlaGlyLeuProLys  542

1807  CAGGGATGGGGCCA.........CCTGGGACAGCAGGAAGCCACTATCCA  1767
 300  Gln..MetGlyProArgGlyLeuProGlyGluArgGly.........Pro  313

1031  AAGCCAAGGCGGCCCGGGCCATGGAACGGTGGGGCGACGAGGAGGGGCC  982
 542  SerProGlyThrGlyProAspGly.........LysThrGlyPr   555

1766  GGATGGCGAGGTCCAGGCAGATGCCCCGGCCCGGAA...............  1731
 314  GlyAlaProGlyAlaArgGlyAlaAlaGlyAsnAspGlyAlaThrGlyAl   330

 981  GACAGCCCTTCTGCTGGCTCGGTGGGGCCCACGCG.................  948
 555  oPro....................GlyProAlaGlyAsnAspGlyArgGlyP   565

1730  ......CCACCCTGCCTCGGTGGGCTCACCCACCACCA.............  1696
 330  aAlaGlyProProGlyProThrGlyProAlaGlyProProGlyPheProG  347

 947  ..CTGCCTCCTCAGCCACCA.................................  930
 565  roGlyProProGlyProGlyAlaArgGlyGlnAlaGlyValMetGly  581
```

Wed Sep 29 14:26:57 1999

```
929 ..................GCAGTGTGGCTGCTACGCAGGTGAGGAAGATG 898
                     ||:::                                    |:::    |||||  ||  :::!||||:::
582 PheProGlyProLysGlyGlyAlaAla..............GlyGluPr 592    854 yAsnValGlyAlaProGlyAlaLysGlyAla.............ArgG 866

897 AGGTGAGGCAGGCCAAAGAGGCACT.....CCTCCTGGGCTGCCCAGGT. 855    222 GCTCTGCT.....CCAGAAGCTGCGGCCTCTCCTTCCTTGCTGCCGCCAA 179
    |||  ||||||:!:||||||  |||||||||:!: |||                   |||||||   |||  ||||||||  |||||||||   |||   ::!:
592 oGlyLysAlaGlyCluArgGlyValProGlyProGlyAlaValGlyP 609       866 lySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArg 882

854 ..............................AGG 852                      178 CTGCCTAGGAATCAGCCAGGCGCCATTCTGCCAGCCCTTTGGTGCCGG 129
                                  |||                           :::       ||||:::||:::||
609 roAlaGlyLysAspGlyGluAlaGlnGlyProProGlyProAla 625          883 ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGl 899

851 GGGCCAGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA.........    813     128 TCCAGCTTCTCAGCCCATGCTCAACACCTGCTGCTGTGGG.....GCAC 85
    |||||||:::                            |||  |||||           |||||||:::||                            |||:!:
626 GlyProAlaGlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPh 642     899 yProAlaGlyLysGlu.....................GlyGlyLysGlyP 909

812 ..........GGTAGCCCAGGCAGCCCCAAGACTGATCATGAAGGC 776        84  CTCAGTGGGGACACGTCTCATCACTCAGATCCTGGCCGA............ 46
          |||      |||     ::: :::                                |:::  !|:::||||     :::  |||    |||||:::
642 eGlnGlyLeuProGlyProAlaGlyProProGlyGluAlaGlyLysProG 659     909 roArg.GlyGluThrGlyProAlaGlyArgProGlyGluValGlyProPr 925

775 ATAGACAGAGTAGGCCTGGCACAGTGT...CCGGGTCCC......GGA 735      45  .....................GGCGCGGCGTGTCACCCGGAGCC 22
    :::|||:! |||||||  ||| ||| |||||||||     :::                                   |||     |||||||||||
659 lyGluGlnGlyValProGlyAspLeuGlyValAlaProGlyProSerGlyAla 675  925 oGlyProProGlyProAlaGlyCluLysGlySerProGlyAla 939

734 AGAGGTCAGAGACAGGAGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAG 685
    ||||||:::||   |||   ::!: |||||||||   ::!:!  |||
676 ArgGlyGluArg...GlyPheProGlyGluArgGlyValGlnGlyProPr 691     seq_name: SwissProt_37:CA15_HUMAN

684 AAGTCCAGGCCCCACGCCCAGAGTGAGCAGTGCCA.............. 648     seq_documentation_block:
    ::!:||||||::!:!||||    ::: :::||||||                     ID   CA15_HUMAN      STANDARD;       PRT;   1838 AA.
691 oGlyProAlaGlyProArgGlyAlaAsnGlyAlaProGlyAsnAspProGlyA 708 AC   P20908;
                                                             DT   01-FEB-1991 (REL. 17, CREATED)
647 ...............GCTCCAGGGGCCTGGGATCCGGGCACAGCAGCC 615     DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
                   |||||||||||       |||:::  :::                DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
708 laLysGlyAspAlaGlyAlaProGlyAla.......ProGlySerGlnGly 722   DE   PROCOLLAGEN ALPHA 1(V) CHAIN PRECURSOR.
                                                             GN   COL5A1.
614 CTGCTAGCCAGCCGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCC 565    OS   HOMO SAPIENS (HUMAN).
    :::     :::    ||||||||||||:::|||||||||||                 OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
723 AlaProGlyLeuGlnGlyMetProGlyGluArgGlyValAlaAlaGlyLeuPr 739 OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
                                                             RN   [1]
564 AAGGACAGTGCCCAGATGAAGGGCCGGCGGGCGGCCATAGCGTCCACGCCA 515    RP   SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.
    |           |||:::          ::: |||||||||:::          RX   MEDLINE; 91302336.
739 oGly......ProLysGlyAspArgGlyAspAlaGlyProLysGlyAlaA 754    RA   TAKAHARA K., SETO Y., OKASAWA K., OKAMOTO N., NODA A., YAOI Y.,
                                                             RA   KATO I.;
514 GTGGTCACTGCTGAGCCTAGGAGCC...GGACACAGA.............. 480   RT   "Complete primary structure of human collagen alpha 1 (V) chain.";
    ::!||       |||  |||     |||                               RL   J. BIOL. CHEM. 266:13124-13129(1991).
754 spGly.......SerProGlyLysAspGlyValArgGlyLeuThrGly 767      RN   [2]
                                                             RP   SEQUENCE OF 621-822.
479 ..........CCAGGCCCAGCACTGGACCAATGCCCA.............. 453   RC   TISSUE=CHORIOAMNIOTIC MEMBRANE;
          ||||||||   ||||||||::!||                              RX   MEDLINE; 89227189.
768 ProIleGlyProGlyProAla.......GlyAlaProGlyAspLysGl 782      RA   SEYER J.M., KANG A.H.;
                                                             RT   "Covalent structure of collagen: amino acid sequence of three
452 ..........................................GCAC 449       RT   cyanogen bromide-derived peptides from human alpha 1(V) collagen
                                              |||               RT   chain.";
782 yGluSerGlyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlaP 799    RL   ARCH. BIOCHEM. BIOPHYS. 271:120-129(1989).
                                                             RN   [3]
448 CATGGTCATGAACTTCTCCTCTACCCCACTTCCAGCAGCAGAGGCGGCA 399    RP   SEQUENCE OF 823-950, AND HEPARIN-BINDING.
    ||  :::      ||||||||||||  |||||:::  |||:::                RC   TISSUE=CHORIOAMNIOTIC MEMBRANE;
799 roGlyAspArgGlyGluProGlyProProGlyProAlaGlyPheAlaGly 815    RX   MEDLINE; 90366601.
                                                             RA   YAOI Y., HASHIMOTO K., KOITABASHI H., TAKAHARA K., ITO M., KATO I.;
398 CATAGGTGATGCTGCGGCCAAACAACACCTCCAGGC............CA 361    RA   MANN K.;
            ||||||||                               |||          RT   "Isolation of the alpha 3-chain of human type V collagen and
816 ...............................ProProGlyAlaAspGlyGlnPr 823 RT   characterization by partial sequencing.";
                                                             RL   BIOL. CHEM. HOPPE-SEYLER 373:69-75(1992).
360 AAGGTTAGCAGGTTGACCAGCAAGAGCTGGGCTTTCCGGTGCCAGCAG 311      RN   [5]
    |||   :::        ||||:::  |||:!:||                          RP   SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.
823 oGlyAlaLysGlyGluProGlyAspAlaGlyAlaLysGlyAspAla.... 838   RC   TISSUE=CHORIOAMNIOTIC MEMBRANE;
                                                             RX   MEDLINE; 94237164.
310 GCGGGTCACCCACAGCCTCTGGACCATAGTGGGCCA............GG 273   RA   MORADI-AMELI M., ROUSSEAU J.C., KLEMAN J.P., CHAMPLIAUD M.F.,
    |||   ||                           ::!:||||               RA   BOUTILLON M.M., BERNILLON J., WALLACH J.M., VAN DER REST M.;
839 ..GlyProProGlyGluProGlyPro.AlaGlyProProGlyProIleGl 854    RT   "Diversity in the processing events at the N-terminus of type-V

272 CGGGTAGGGCTCAGGGGGCCGGTTCAGGCACTCCAGAACTGCTTCGTCCG 223
```

```
RT       collagen.";
RL       EUR. J. BIOCHEM. 221:987-995(1994).
RN       [6]
RP       VARIANT EDS1 SER-1639.
RX       MEDLINE; 97195540.
RA       DE PAEPE A., NUYTINCK L., HAUSSER I., ANTON-LAMPRECHT I.,
RA       NAEYAERT J.-M.;
RT       "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos
RT       syndromes I and II.";
RL       AM. J. HUM. GENET. 60:547-554(1997).
CC   -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
CC       COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
CC       TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
CC       MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
CC       ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- PTM: 40% OF TYROSINES IN THE PRO-ALPHA 1(V) CHAIN ARE SULFATED.
CC   -!- DISEASE: DEFECTS IN COL5A1 ARE A CAUSE OFEHLERS-DANLOS SYNDROME,
CC       TYPE I (EDS1), A DISEASE CHARACTERIZED BY LOOSE-JOINTEDNESS AND
CC       FRAGILE, VELVETY, STRETCHABLE, BRUISABLE SKIN THAT HEALS WITH
CC       PECULIAR 'CIGARETTE-PAPER' SCARS.
CC   -!- SIMILARITY: HIGH, TO ALPHA 3(V) AND ALPHA 1(XI) CHAINS.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; D90279; G219510; -.
DR   PIR; S03978; S03978; .
DR   PIR; S11303; S11303.
DR   PIR; S16024; S16024.
DR   MIM; 120215; -.
DR   MIM; 130000; -.
DR   MIM; 130010; -.
KW   EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW   COLLAGEN; SIGNAL; HEPARIN-BINDING; SULFATATION; DISEASE MUTATION.
FT   SIGNAL        1     37       POTENTIAL.
FT   CHAIN        38   1605       COLLAGEN ALPHA 1(V) CHAIN.
FT   DOMAIN      444    558       NONHELICAL REGION.
FT   DOMAIN      559   1570       INTERRUPTED COLLAGENOUS REGION.
FT   DOMAIN     1571   1605       TRIPLE-HELICAL REGION.
FT   PROPEP     1606   1838       NONHELICAL REGION.
FT                                CARBOXYL-TERMINAL PROPEPTIDE.
FT   MOD_RES     570    570       HYDROXYLATION.
FT   MOD_RES     576    576       HYDROXYLATION.
FT   MOD_RES     621    621       HYDROXYLATION.
FT   MOD_RES     627    627       HYDROXYLATION.
FT   MOD_RES     639    639       HYDROXYLATION.
FT   MOD_RES     642    642       HYDROXYLATION.
FT   MOD_RES     648    648       HYDROXYLATION.
FT   MOD_RES     654    654       HYDROXYLATION.
FT   MOD_RES     657    657       HYDROXYLATION.
FT   MOD_RES     675    675       HYDROXYLATION.
FT   MOD_RES     678    678       HYDROXYLATION.
FT   MOD_RES     680    680       HYDROXYLATION.
FT   MOD_RES     686    686       HYDROXYLATION.
FT   MOD_RES     690    690       HYDROXYLATION.
FT   MOD_RES     696    696       HYDROXYLATION.
FT   MOD_RES     705    705       HYDROXYLATION.
FT   MOD_RES     708    708       HYDROXYLATION.
FT   MOD_RES     717    717       HYDROXYLATION.
FT   MOD_RES     720    720       HYDROXYLATION.
FT   MOD_RES     726    726       HYDROXYLATION.
FT   MOD_RES     732    732       HYDROXYLATION.
FT   MOD_RES     744    744       HYDROXYLATION.
FT   MOD_RES     750    750       HYDROXYLATION.
FT   MOD_RES     756    756       HYDROXYLATION.
FT   MOD_RES     762    762       HYDROXYLATION.
FT   MOD_RES     765    765       HYDROXYLATION.
FT   MOD_RES     771    771       HYDROXYLATION.
FT   MOD_RES     774    774       HYDROXYLATION.
FT   MOD_RES     780    780       HYDROXYLATION.
FT   MOD_RES     789    789       HYDROXYLATION.
FT   MOD_RES     795    795       HYDROXYLATION.
FT   MOD_RES     804    804       HYDROXYLATION.
FT   MOD_RES     807    807       HYDROXYLATION.
FT   MOD_RES     810    810       HYDROXYLATION.
FT   MOD_RES     816    816       HYDROXYLATION.
FT   MOD_RES     819    819       HYDROXYLATION.
FT   MOD_RES     834    834       HYDROXYLATION.
FT   MOD_RES     846    846       HYDROXYLATION.
FT   MOD_RES     861    861       HYDROXYLATION.
FT   MOD_RES     864    864       HYDROXYLATION.
FT   MOD_RES     870    870       HYDROXYLATION.
FT   MOD_RES     873    873       HYDROXYLATION.
FT   MOD_RES     876    876       HYDROXYLATION.
FT   MOD_RES     882    882       HYDROXYLATION.
FT   MOD_RES     888    888       HYDROXYLATION.
FT   MOD_RES     891    891       HYDROXYLATION.
FT   MOD_RES     897    897       HYDROXYLATION.
FT   MOD_RES     903    903       HYDROXYLATION.
FT   MOD_RES     906    906       HYDROXYLATION.
FT   MOD_RES     930    930       HYDROXYLATION.
FT   MOD_RES     945    945       HYDROXYLATION.
FT   MOD_RES    1017   1017       HYDROXYLATION.
FT   MOD_RES    1020   1020       HYDROXYLATION.
FT   MOD_RES    1023   1023       HYDROXYLATION.
FT   MOD_RES    1029   1029       HYDROXYLATION.
FT   MOD_RES    1221   1221       HYDROXYLATION.
FT   MOD_RES    1224   1224       HYDROXYLATION.
FT   MOD_RES    1467   1467       HYDROXYLATION.
FT   MOD_RES    1470   1470       HYDROXYLATION.
FT   VARIANT    1639   1639       C -> S (IN EDS1).
FT   CONFLICT    641    641       E -> G (IN REF. 2).
FT   CONFLICT    650    650       P -> L (IN REF. 2).
FT   CONFLICT    663    663       R -> E (IN REF. 2).
FT   CONFLICT    668    668       E -> Q (IN REF. 2).
FT   CONFLICT    677    677       K -> Q (IN REF. 2).
FT   CONFLICT    684    684       L -> Q (IN REF. 2).
FT   CONFLICT    692    699       PPGPGVT -> VTGEPGAP (IN REF. 2).
FT   CONFLICT    727    727       G -> Q (IN REF. 2).
FT   CONFLICT    741    741       P -> L (IN REF. 2).
FT   CONFLICT    747    747       L -> Q (IN REF. 2).
FT   CONFLICT    753    753       P -> A (IN REF. 2).
FT   CONFLICT    759    759       D -> N (IN REF. 2).
FT   CONFLICT    776    777       GQ -> QK (IN REF. 2).
FT   CONFLICT    849    855       GGPNGDP -> IGPPGPR (IN REF. 3).
FT   CONFLICT    894    894       N -> D (IN REF. 3).
SQ   SEQUENCE   1838 AA;  183616 MW;  1D5EBBFA CRC32;


alignment_scores:
           Quality:  337.50                     Length:   1198
           Ratio:    0.692                       Gaps:     79
Percent Similarity:  40.735      Percent Identity:  25.876

alignment_block:
US-09-030-606-110/rev x CA15_HUMAN

Align seg 1/1   to: CA15_HUMAN    from:   1  to: 1838

3068 GGTCCTGCAGTAGCTCCAAACACAGGTTGTGGAGCTGGTGGGAAAGTTGG 3019
     |||||:||::::::      :::    |||     ||
431  GlyProGlyMetProAlaAsnGlnAspThrIleTyrGluGlyIleGlyGl 447

3018 GGGTAGGGGAAAGTTGGGG...GTAGGGGAAATTTGGGCAGTGCCTTCA 2972
     ||||||:::::  |||:::::  |||::|
447  yProArgGlyGluLysGlyGlnLysGlyGlu........ProAlaIleI 461
```

```
2971 TCAGGCCAGTCCTAGAGAGGAGTAGAGGG...............          2943
     ||:::|||  ::    :: :||:||||
 461 leGluProGlyMet...LeuIleGluGlyProProGlyProGluProGlu    476

2942 ...........AGTGGAAGTGGGGGAACCAGGC                      2920
     |||             |||::|: |||
 477 AlaGlyLeuProProGlyProThrMetGlyProThrGlyGlnValGl       493

2919 TGGGCCAAGAGAGAAGAGGGGTGGTTAGGAAGCCGTTGAGACCTGAAGCC.   2871
     ;  |||  ||||:::|||  ||| ||| |||
 493 yAspProGlyGluArgGlyProProGlyArgProGlyLeuProGlyAlaA    510

2870 ..........CCACCCCTACCTTCCTTCAACACCCTAACCTTGGG          2835
     |||||:::|||:||::: |||
 510 spGlyLeuProGlyProProGlyThrMetLeuMetLeuPro........     523

2834 TAACAGCATTTGGAATTATCATTTGGGATGATGAGTAGAATTCCAAGGTCCT  2785
     ::: |||||||
 524 ...........PheArgPheGly.........................      527

2784 GGGTTAGGCCATTTTGGGGGGCCAGA..............CCCCAGGAGA    2750
     |||   ::: ||||::||::       |||||:
 528 .GlyGlyAspAlaGlySerLysGlyProMetValSerAlaGlnGluS      544

2749 AGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGACCT   2700
     ::::::  |||:::||           ||||||||||
 544 erGlnAlaGlnAlaIleLeuGlnGlnAlaArgLeuAlaLeuArgGlyPro   560

2699 GATTGTTGGGGATCCCCACCCTACCCCAAATATTAGACACCAACACAGAA   2650
     |||           |||
 561 Ala......GlyProMetGlyLeu.........................     566

2649 AAGCTAGCAATGGATCCCTTCTACTTTGTTAAATAAATAAGTTAAATAT    2600
     |||||||||||||  ::: |
 566 ................................................     566

2599 TTAAATGCCTGTCTCTGTGATGGCAACAGAAGGACCAACAGGCCACAT    2550
     ||:::||||||
 567 ..................................ThrGlyArgProGly  571

2549 CCTGATAAAAGTAAGAGAGGGGGTGGATCAGCAAAAAGACAGTGCTGTGG  2500
     |||            ||:::|||
 572 ProValGlyPro...ProGlySerGly...................Gl  580

2499 GCTGAGGGGACCTGGTTCTTGTGTGTTGCCCTCAGGACTCTTCCCCTAC   2450
     :::::||| :||          |||
 580 yLeuLysGlyGluProGlyAspValGlyPro..................   590

2449 AAATAAGTCATATGTTCAAATCCATGGAGGAGTGTTTCATTCCTAGAAAC   2400
     :::||        |||   :||
 591 ..........GlnGlyProArgGlyValGlnGlyProProGly.......   601

2399 TCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTAAGGGCTTAGAG     2350
     :::||        ||||| :: ||
 602 ProAlaGlyLysProGlyArgArgGlyArgGlyAlaGlySerAspGlyAlaAr 618

2349 ATGGGAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAACCCTTCTCTA  2300
     |  ||| ||||||    ::::||||::::
 618 g.GlyMetProGlyGln.........ThrGlyProLysGly...AspArg  630

2299 GGTGTGTCTCAACTAGGAGCTAGCTGTTAACCCTGAGCCTGGGTAATCC    2250
     ::: |||    ||||||||||||
 631 GlyPheAspGlyLeuAlaGly.........................Le    638

2249 ACCTGCAGAGTCCCCGCATTCCAGTGCATGGAGCCCCTTCTGCCTCCCTG  2200
     |||:::|||:::::  ||| ::::||||||||||  |||
 638 uProGlyGlyGluLysGlyHisArgGlyAspProGlySerGlyProPro.   654

2199 TATAAGTCCAGACTGAAACCCCCTTGGAAGGCCTCCAGTCAGGCAGCCCT  2150
     :::  |||
 655 ..........GlyProPro.........................        657

2149 AGAGACTGGGGAGAGG                                    2124
                                          AGAGGGAC
```

```
     |||  |||||||||                                      674
 658 GlyAspAspGlyGlyAspArgGlyAspAspGlyGlyGluValGlyProArgGlyIle

2123 GCCCCAGCCCCAGCTGTGCAG......CTACGCACCTCAGCAGCACAGG   2080
     ||| |||                ::::::
 674 uProGlyLysProGlyProArgGlyLeuLeuGlyProLysGlyProProG  691

2079 GTGGCAGCAGAGAGCAGCCACAATTACTTTGGCAGCAACAACAGAAACTGGCGGCCCA 2030
     ||            ||| ::|||                    ||||
 691 lyProProGlyProPro........Gly.ValThrGlyMetAspGlyG1  704

2029 GCCCGGCAGCCCCATGGGGCTAACAGGA...GCGGGGAGCTGGGACCCAG 1983
     ||||||| |||  |||::::|||         |||
 704 nProGly...ProLysGlyAsnValGlyProGlnGlyGluProGlyProP  720

1982 TGAGGCAGGCCCTCCACCCCATGTCTGGGAGTTTTCTACGCTGAGTAT    1933
     ||                                  ::::|||
 720 roGlyGlnGlnGlyAsnProGly...............AlaGlnGly    730

1932 TTGGCCAAGTGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGC  1883
     ||| ||:::        ||  |||   ::::::
 731 Leu.ProGlyProGlnGlyAlaAlaIleGlyProProGlyGluLysGlyProL 747

1882 GACCAGACACCCAGGCCTGCGGCAGCACACCCATATAGGCAGTGACAGACTGGC 1833
     ||||||||       ::  |||            :::
 747 euGlyLysProGlyLeuProGlyMetProGlyAlaAspGlyProProGly 763

1832 TGAGCTGGACAATGGAGCCCATAAACAGGGATGGGGCGCCACCTGGG...AC 1786
     ||||||||||||||||||
 764 HisProGlyLys...............GluGlyProProGlyGluLyU    774

1785 AGCAGGAAGGCACTATCCAGGATGGGCAGGTCCAGGATGGCCCCGGCC   1736
     ::::|||::  |||||||    :::||||||
 774 sGlyGlyGlnGlyProGlyProGlnGlyProIleGlyProIleGlyTyrProGlyP 791

1735 CGGAA...........                                   1731
     ||:::
 791 roArgGlyValLysGlyValAlaAspGlyLeuIleArgGlyLeuLysGlyThrLys 807

1731 ...............                                    1731

 808 GlyGluLysGlyGluAspGlyPheProGlyPheLysGlyAspMetGlyIl 824

1730 ...........CCACCCTGGCCTCCGGTGGGCTCACC               1706
                ||||||||||  |||:::
 824 eLysGlyAspArgGlyGluIleGlyProProGlyProArg..GlyGlu.. 839

1705 CACCACCACAGTACGGAGAGCAATCACAGGCAGGAGGCCCCGAGAGCGCGG 1656
     :::        :::||||::||::::
 840 ...............AspGlyProGluGlyProLysGlyArgG         849

1655 GTGGAGGTGGAGCAGGCAGGCCACTGCCTCCAGCACCACGTGTCCATTAGG 1606
     |||||  ::::: |||  :::|||       :::|||
 849 lyGlyProAsnGlyAspProGlyProLeuGlyProProGlyGluLysGly  865

1605 AAG...GGAGCTCCAGGCTTAGGG.......CCTGGCAGGAAGCTGGTCAT 1565
     |||  :::::|||||||||       ||||||||||||||
 866 LysLeuGlyValProGlyLeuProGlyThrProGlyTyrProGlyArgGln. 879

1564 CAGGCTGTCCTCACTGCTAGCACCTCCAGTGTCCCCTCGGTATT......  1521
     :::             :::|||  ::::|||
 880 ..............GlyProLysGlySer.IleGlyPheProGly       889

1520 TGGGCAGGAAACAACCTGCTTCTCCCGGTGGTAGAGGGAGGCCAGTGTGTAG 1471
     ::: ||:::  |||:::::  |||||||||||| :::
 890 PheProGlyAlaAsnGlyLysGlyGluLysGlyGlyArgGlyThrPro.   903

1470 GGCAGGATCTGCAGGGGCTGAGAAGGTGAACCCGGTGAGGGGCGCTGAAGC 1421
     ||:::::  |||
 904 ...GlyLysProGly.............................         907

1420 TGTCACCACGGCCCACACTGTGGACAACAGGCATGTGGCACCGGCAGCCACAG 1371
                                               |||:::
```

```
605  AGCCGGCCCTTGGGATGAG.....................AAAGAGG  580
          :::!!!!!!       !!!
1148 oValGlyProProGlyGluAspGlyLysGlyGluIleGlyGluProG  1165

579  CTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAAGGGCCGGCGCGCC  530
          ::!!!:!!       :::!!!!!!!!    !!!!!!!!!!!! !!
1165 lyGlnLysGlySerLysGlyAspLysGlyGluGlnGlyProProGly.Pr  1181

529  ATAGCGTCCACCGCCAGTGTCACTGGCTGAGCCTAGGA.........GCG  489
     |    !!!!! !!!!     ::::!!!!!!:::!!! !!!      !!!
1181 oThrGlyPro...GlnGlyProIleGlyGlnPro.GlyProSerGlyAla  1196

488  GGACACAGACCAGGCCCAGCACTGGACCAA...................  459
          :::!!!!!!!!!!!               ::!!!
1197 AspGlyGluProGlyProArgGlyGlnGlnGlyLeuPheGlyGlnLysGl  1213

458  ...........................TGCCCAGCACCATGGTCATGAACTTCT....  432
                                  !!!::!!      !!!  !!!
1213 yAspGluGlyProArgGlyPheProGlyProProGlyProValGlyLeuG  1230

431  .........CCTCTACCCCACTTCCAGCACGCAGAGGCGGCACATAGGTGA  390
                   !!! !!!!!        ::!!!!!!!:::
1230 lnGlyLeuProGlyProGlyLeuLysGlyGluThrGlyAspValGly  1246

389  TGCCTGCGCCAAACACACCTCCAGGCCAAAGG...TTAGCAGTTGACC  343
          ::!!!   !!!!!!!!!!   !!           !!!::!!!  !!
1247 GlnMetGlyProGlyProProGlyProArgGlyProSerGlyAlaPr  1263

342  AGCAAGAGCTGGGCTTTCCGGTGCCGACCAGGC...GGCTCACCCACAG  296
          !!!::!!!    :!!:!!!!!     :::!!      !!!!!!!!!
1263 oGlyAlaAspGlyProGlnGlyProGlyIleGlyAsnProGlyA  1280

295  CCTCTGGACCAATAGTGGG.............CCAGGCGGGTAGGGCTCAGG  258
     !!::!!!!!!!:::!!!                  !!!
1280 laValGly.GluLysGlyGluAlaGlyLeuProGlyProSe  1296

257  GGGCGGTTCAGGCACTCCA................GAACTGCTTCGTCTCGGC.  221
     :!!!!!!!!!!!  !!!                   !!!
1296 rGlyArgSerGlyProProGlyProGlyLysGlyGluLysGlyGlyG  1313

220  ..TCTGCTCCAGAAGCTGCGGGCCTCTCCTCCTGCTCCGCGC..........  182
           ::!!!!!!!!!::!!!!!!!          :::
1313 luSerGlyProSerGlyAlaAlaGlyProProGlyProLysGlyProPro  1329

181  ......CAACTGCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTT  138
            ::!!!!!:::!!!!!!!!!!                  !!!
1330 GlyAspAspGlyProLysGlySerProGly...............ProVa  1341

137  TGGT.........GCCGGTCAGCTTCTCAGCCCATGCTCAACACCT  100
     !!!        !!!!!!!  ::::!!!             ::: !!!
1341 lGlyPheProGlyAspProGlyProGlyGluPro.....Gly.Pro  1355

99   GCTGCTGTGGGGCACCTCAGTGGGACAACGTCTCATCACTCAGATCCTGG  50
     !!!::!       :::!!!!!            ::!      :::!!!!!!!!
1356 AlaGlyGlnAspGlyProProGlyAspProLysGlyAspAspGlyGlyProGl  1372

49   CCGAGGCCGCGCCGTGTCACCCGGA  25
     !:::!      !!!!!!!!!!!!!
1372 yGlnThrGly......SerProGly  1378
```

seq_name: SwissProt_37:CAIF_HUMAN

```
908  ..............................ProArg  909

1370 GGAAAGCTGCCACACTGGCCAAATAGACTGCTGCTCGAGTGCCGAATCGCTGC  1321
     !!!::!!                            !!!!!
910  GlyGln.............................ArgGl  913

1320 ACCAGCCGGTCCATGACCAGCAGAGA.....................AGAC  1292
     !!!!!:::!!!!!!!                                !!!!
913  yProThrGlyProArgGlyGluArgGlyArgGlyIleThrGlyIleLysP  930

1291 CAGGG...AGATGGCGCACTGCAGGAACAGCCCA..GGCTGCCCATCC  1248
     !!!!!   !!::!!!!!!:::!!!   !!!!!     !!! !!!
930  roGlyProLysGlyAsnSerGlyGlyAspGlyProAlaGlyProProGly  946

1247 GAACGGCCTTCATCATACTGTCTCCGGCGTCCGGTGCCCGGCTCAGCTCTG  1198
     !!!!!                    !!!!!!!!!!!
947  GluArg.............GlyProAsnGlyProGlnGlyProTh  957

1197 GGCACGCCCTGTACAGCCCCTCCGCCCACGAAATCCGTGTAAAAACAGCGT  1148
     :::  !!!!!!    !!!    !!!           !!!  ::: !!!!
957  rGlyPheProGlyLysGlyProProGlyProProGlyLysAspGly.  973

1147 GAAGGTCATGAGTGCCA.......................  1131
                   ::!!!
974  ...........LeuProGlyHisProGlyGlnArgGlyGluThrGlyGlyPhe  986

1130 TCCAGCTGCACAGCTCAGCCACGAAGAGCCGGCGCAGGGTCGGGGCATG  1081
     :::!!!!:::   !!!    !!!:::!!!!!!!!     !!!!!!!!!!!
987  GlnGlyLysThrGlyProProGlyProGlyVal...ValGlyProGl  1002

1080 CGGCAGCACA.......GCTGGTGCGAGCCGGGAAGCAGGCGCCCAGGTT  1037
     !!!::!!!           !!!!!!       !!!!!!:::!!!!!!!!!
1002 nGlyProThrGlyThrGlyProMetGlyGlnArgGlyHisProGlyP  1019

1036 CCCGGAAAGCCAAGCGGG...............CCCGGCATGGACAGCAGTGGG  999
     !!!  !!                         !!!!!!:::::
1019 roProGlyProGlyGluGlnGlyLeuProGlyLeuAlaGlyLysLysGlu  1035

998  GCGACAAGGAGGAGGGGCCGACAGCCCTTCTGCTGGGGGCCCAGC  949
     :!!!!!::::!!!!!!                      !!!!
1036 GlyThrThrGlyGlyAspPro...............GlyProAl  1044

948  G...CTGCCCCTCAG......CCACCAGCAGTGTGGCTGCTACGGCAGGT  908
     !   !!!!!!!        !!!!!!
1044 aGlyLeuProGlyLysGluAspGlyGlyPro.......................  1053

907  GAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCACT........................  873
     !!!          :::!!!!!!!::
1054 ..GlyLeuArgGlyPheProGlyAspProArgGlyLeuProGlyProValGly  1069

872  ...............CCTCCTGGGTGCCCAGGT......  855
                    !!!!!!!
1070 AlaLeuGlyLeuLysGlyAsnGluGlyProGlyProProGlyProProGlyProAl  1086

854  .AGGGGGCCAGGGCACTGGTGTCCCAGTCAATGGCAGGCAGGAGGTAGCC  806
     :::!!!!!!:::   !!!::!
1086 aGlySerProGlyGluArgGlyProAla......GlyAlaAlaGlyProI  1101

805  CAGGCAGCCCCAAGACTGATCATGAGGCATAGACAGATAGGCCTGGC  756
     !!!  !!!!!!!!!!!          ::!!!!!!!!!!
1101 leGlyIlePro...........................GlyArgProGly  1108

755  GACAGTGGTCCGGGTCCGGGACCGGGAAGAGCCAGGAGCCAGGGCCTCCAGTGGA  706
     !:::!!!!!!                    !!!!!!
1109 ProGlnGlyPro....................ProGlyProAlaGlyGl  1118

705  GTGAAGCACACTGGCCACAGAAGTCAGCACGCCCACAGCCCCAGGATGAG  656
     !:::!!!!!!:::::!!!!!:::!!!   :::!!!     !!!!!!
1118 uLysGlyAlaProGlyGluLysGlyProGlnGlyProAlaGlyArgAsp  1134

655  CAGTGCCAGCTCAGGGGCTCCAGGGATCCGGGCCACAGCAGCCCCTGCTAGCC  606
     !!!!!!!!!!!        :::!!!!!!                 !!!
1135 ......GlyLeuGlnGlyPro...ValGlyLeuProGlyProAlaGlyPr  1148
```

```
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92335339.
RA   PAN T.C., ZHANG R.Z., MATTEI M.-G., TIMPL R., CHU M.-L.;
RT   "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:6565-6569(1992).
RN   [2]
RP   SEQUENCE OF 418-1603 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 93203161.
RA   YAMAGUCHI N., KIMURA S., MCBRIDE O.W., HORI H., YAMADA Y.,
RA   KANAMORI T., YAMAKOSHI H., NAGAI Y.;
RT   "Molecular cloning and partial characterization of a novel collagen
RT   chain, alpha 1(XVI), consisting of repetitive collagenous domains and
RT   cysteine-containing non-collagenous segments.";
RL   J. BIOCHEM. 112:856-863(1992).
CC   -!- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
CC       THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
CC   -!- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
CC       AMNION, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE
CC       AMNION, IT IS FOUND IN AN ACELLULAR, RELATIVELY DENSE LAYER OF A
CC       COMPLEX NETWORK OF RETICULAR FIBRES. ALSO LOCATED TO A FIBROBLAST
CC       LAYER BENEATH THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION
CC       WITH OTHER TYPES OF COLLAGEN.
CC   -!- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
CC       GESTATION, AND DECREASE AT TERM.
CC   -!- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
CC       TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
CC       DOMAINS (NC10 TO NC1).
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: BELONGS TO THE FAMILY OF FIBRIL-ASSOCIATED COLLAGENS
CC       WITH INTERRUPTED HELICES (FACIT).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by  non-profit institutions as  long  as its content is in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M92642; G180758; -.
DR   EMBL; S57132; G298642; -.
DR   PIR; S23810; S23810.
DR   MIM; 120326; -.
KW   EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; COLLAGEN; HYDROXYLATION;
KW   REPEAT; SIGNAL.
FT   SIGNAL          1     21       POTENTIAL.
FT   CHAIN          22   1603       COLLAGEN ALPHA 1(XVI) CHAIN.
FT   DOMAIN         22    374       NONHELICAL REGION 10 (NC10).
FT   DOMAIN        375    505       TRIPLE-HELICAL REGION 9 (COL9)
FT                                  WITH 3 IMPERFECTIONS.
FT   DOMAIN        506    520       NONHELICAL REGION 9 (NC9).
FT   DOMAIN        521    554       TRIPLE-HELICAL REGION 8 (COL8)
FT                                  WITH 1 IMPERFECTION.
FT   DOMAIN        555    571       NONHELICAL REGION 8 (NC8).
FT   DOMAIN        572    630       TRIPLE-HELICAL REGION 7 (COL7)
FT                                  WITH 1 IMPERFECTION.
FT   DOMAIN        631    651       NONHELICAL REGION 7 (NC7).
FT   DOMAIN        652    722       TRIPLE-HELICAL REGION 6 (COL6)
FT                                  WITH 1 IMPERFECTION.
FT   DOMAIN        723    737       NONHELICAL REGION 6 (NC6).
FT   DOMAIN        738    875       TRIPLE-HELICAL REGION 5 (COL5)
FT                                  WITH 3 IMPERFECTIONS.
FT   DOMAIN        876    886       NONHELICAL REGION 5 (NC5).
FT   DOMAIN        887    938       TRIPLE-HELICAL REGION 4 (COL4)
FT                                  WITH 2 IMPERFECTIONS.
FT   DOMAIN        939    972       NONHELICAL REGION 4 (NC4).
FT   DOMAIN        973    987       TRIPLE-HELICAL REGION 3 (COL3).
FT   DOMAIN        988   1010       NONHELICAL REGION 3 (NC3).
FT   DOMAIN       1011   1432       TRIPLE-HELICAL REGION 2 (COL2)
FT   DOMAIN       1433   1471       NONHELICAL REGION 2 (NC2).
FT   DOMAIN       1472   1577       TRIPLE-HELICAL REGION 1 (COL1)
FT                                  WITH 2 IMPERFECTIONS.
FT   DOMAIN       1578   1603       NONHELICAL REGION 1 (NC1).
FT   CONFLICT      418    420       RDA -> GGR (IN REF. 2).
FT   CONFLICT      537    537       R -> P (IN REF. 2).
FT   CONFLICT     1160   1160       T -> P (IN REF. 2).
FT   CONFLICT     1163   1163       T -> P (IN REF. 2).
FT   CONFLICT     1165   1165       S -> P (IN REF. 2).
SQ   SEQUENCE   1603 AA;  157692 MW;  4BB01544 CRC32;
```

alignment_scores:
```
             Quality:   330.00       Length:   1262
               Ratio:     0.643         Gaps:     81
Percent Similarity:    40.650   Percent Identity:   25.040
```

alignment_block:
US-09-030-606-110/rev x CAIF_HUMAN

Align seg 1/1 to: CAIF_HUMAN from: 1 to: 1603

```
3077 TGTGCTTCTGGTCCT...........................GCAGTAGCTCC 3052
     |||||||::::|||||
 372 CysAlaGluGlyProLysGlyGluLysGlyGluSerGlyAlaLeuGlyPr 388

3051 AAACAGGGTTGTGAGCTGGTGGGGGAAGTTGGGGGTAGGGGAAAGTTGG 3002
     ::: ||:::          ||  :::|||||:   ::|||||: 
 388 oSerGlyLeuProGlySerThrGlyGluLysGlyGlnLysGlyGluLysG 405

3001 GGGTAGGGGAAATTTGGCAGTGCCTTCATCAGCCCAGTCCTAGAGAGA 2952
     ||   |||   |||  |||||||||| ::|||:::  :||::::  ||: 
 405 lyAspGlyIleLysGly.ValProGly.LysProGly...ArgAspA 420

2951 GTAGAGGGGAG...........................TGGAAGTGGGGGAACCA 2923
     ::  |||||||                           |||:::|||| 
 420 laProGlyGluIleCysValIleGlyProLysGlyGlnLysGlyAspPro 436

2922 GGCTGGGCCAAGAGAAGAGAGGGGTGGTTAGGGAGGAGCCGTTGAGACCTGAAG 2873
     ||:::||:  |||       |||  :::|||||    :::   ||| 
 437 GlyPheValGlyProGluGlyLeuAlaGlyGluPro...............Gl 449

2872 CCCCACCCCTACCTTCCTTCAACACCACAGAAAAGCTAGCAATGGATTCCCTCT 2823
     :|||   |||    :|||:::    |||       ||| 
 449 yPro...........ProGlyLeuProGlyPro............... 456

2822 GAATTATCATTGGGATGAGTAGTAGAATTTCCAAGTCCTGGGTAGGCATT 2773
                                    |||||||::||::|||:: 
 457 ........................................ProGlyIleGlyLeu 461

2772 TTGGGGGGCCAGACCCCAGGAGAAGAAGAGAGATTTCGGCAATGATCAGCCCAA 2723
     |||           |||||||:::          ||:::||| 
 462 ProGly......ThrProGlyAsp........................ProGl 469

2722 TGA....CCAGCTATCTCAGGGGACCTGATTGTTGGGGATCCCCCACCCT 2677
     :           |||::::  :::|||||||
 469 yGlyProProGlyProGlyGlyGlyAsp......................... 477

2676 ACCCAAATATTAGACACCAACACAGAAAAGCTAGCAATGGATTCCCTCT 2627
     :||          |||::||::::||:::|||
 478 ................LysGlySerSerGlyIlePro... 484

2626 ACTTTGTTAAATAAATAAGTAAATATTTAAATGCCTGTCTCTGTGAT 2577
                  |||:::||||||
 484 .......................................ProGly 469

2576 GGCAACAGAAGGACCAACAGGCCACATCCTGATAAAAGGTAGAGGGGGG 2527
     ::: |||||||||  |||
 485 .GlyLysGluGlyProGlyGlyLys............ProGlyLysProGlyV 498

2526 TGGATCAGCAAAAGACAGTGCCTGTGGGCTGAGGGGACCTGGTTCTTGTG 2477
     ::||||
     |||||
```

```
1780 GA...AGGCACTATCCAGGATGGCGGAGGTCCAGGCAGA............ 1746
 776 |:  :::    |||||||        ||||||||
     lyValGlnGlyLysGluProGly.....ProProGlyArgGlyValGlnGly 790

1745 ...................................................TGCCC 1741
                                                          ||

 791 ProGlnGlyGluProGlyAlaProGlyLeuProGlyIleGlnGlyLeuPr 807

1740 CGGCCCCGAA...CCACCTGGCCT.........CGGTGGGCTCACCCAC 1703
 807 |||||||||: |||:||||||  |    |||:||||||
     oGlyProArgGlyProThrGluProGlyLyAlaGlnGlys 824

1702 CACCACGTACGGAGACATCACAGGC............................ 1676
 824 |||  |||:|||    ||||||
     erProGlyVallys.GlyAlaThrGlyProValGlyProProGlyAlase 840

1675 ...................................................A 1675

 840 rValSerGlyProProGlyArgAspGlyGlnGlnGlyGlnThrGlyLeuA 857

1674 GAGGC...CCCGCAGACGCGGGTGGAGGTGGGAGGCAGGCCACTGCCTCC 1628
 857 ||||| |||:|:||:::|||| |||:|||  |||
     rgGlyThrProGlyGluLysGlyProArg.GlyGluLysGlyGluProGl 873

1627 AGCACCCACGTCCATTAGGG............................AAGG 1602
 873 ::||||||
     yGluCysSerCysProSerGlnGlyAspProSerGlnGlyMetProG 890

1601 GAGCTCCAGGTTAGGGCCTGGCACCAGGAAGCTGGTCATCAGGCTGTCCTCA 1552
 890 ||||||||||||  ||| ::||||:|||:||
     lyAlaProGlyLeuTrpMetGly.SerSerTrpGlnProGly......... 903

1551 CTGCTAGCACCTCCAGTCCCCCTCGGTATTGGGCAGGAACACCTGCTT 1502
 904 ||||  |||||| ||||:::|   ProGlyProProGlyP 916
     ProGlnGlyProProGlyIle...

1501 CTCCCGGTGGTAGAGGAGGCCAGTGTGTAGGGCAGGATCTCCAGGGCTG 1452
 916 ||||||| |||  ||| ||| |||
     roProGly.ValProGlyLeuGlnGlyVal.........ProGlyAs 928

1451 AGAAGGTGAACCCGGTGAGGGCGGTGAAGCTGTCACCACGGC........ 1409
 928 ::||| ||||||  |||  :::|||:|||
     nAsnGlyLeuProGlyGlnProGlyLeuThrAlaGluLeuGlySerLeuP 945

1408 ...........CACACTGTGGGACAGGCATGTGGCACCGGCAGCCACAGG 1370
 945 ||| ::: ||||||||
     roIleGluGlnHisLeuLeuLysSerIleCysGly.............. 956

1369 GAAAGCTGCCACCTGCCAAATAGACTGCTCGAGTGCCGAATGCGCTGCA 1320
 957 ||||||       :::::||||||
     ........AspCysValGlnGlyValGlyGlnArgAlaHi 965

1319 CCAGCCGTCCATGACCAGGAGAAGACCAGGGAGATGGCGACTGCCGCAGG 1270
 965 ::||| ||||||||||||  :::||||
     sProGlyTyrLeuValGluLysGlyGluLysGlyAspGlnGlyIle..... 980

1269 AACAGCCCCAGGCTGCCCATCCGAACGCCTCATCATAGTGTCTCCGGGC 1220
 981 |||           ||
     ..................ProGlyVal............. ProGly 985

1219 CTCGGTGCCGGCTGCGCTCTGGCACGCTGGGCACCCCTGCTACAGCCCCA 1170
 986 |||  |||||    ::|||  :::|||
     LeuAspAsnCysAlaGlnCysPheLeuSerLeuGluArgProArgAlaGl 1002

1169 CGAAATCCGTGTAAAACAGCGTGAAGTGCATGAGTGCCATCCAGC.TGCA 1121
 1002 :||| ||  :::  ||||||  |||  |||
     uGluAlaArgGlyAspAsnSerGluGly.........AspProGlyCysV 1016

1120 CAGCTCAGCCA.............................CGAAGAGCC 1101
 1016 ::||||:|||
     aGlySerProGlyValGlnGlyLeuProGlyProGlyLeuProGlnArgGly 1032
```

```
 498 aLLysGluLys...............GlyAsp.............. 504

2476 TGTTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAATCC 2427
 505 ...................................Pr 505

2426 CATGGAGGAGTGTTCATCCTAGAAACTCCATGCAAGAGCTACATTAAA 2377
 505 |   |||  ||  |||||::::|:::   ;
     oCysGluvalCysProThrLeuProGluGlyPheGlnAsnPhe......V 520

2376 CGAAGCTGCAGTTAAGGGCTTAGAGATGGGAAACCAGGTGACTGAGTT 2327
 520 ::  |||    |||:||||||||||  |||
     alGlyLeuProGlyLysProGlyProLysGlyGluProGlyAspProVal 536

2326 TATTCA.........................GCTCCCAAAAA 2310
 537 :::                          ||||||
     ArgAlaArgGlyAspProGlyIyIleGlnGlyIyIleLysGlyGluLysGlyGl 553

2309 CCCTTCTCAGGTGTGTCTCAACTAGGAGGCTAGCTGTTAACCCTGAGC. 2261
 553 :|||   |||:::  ||   |||    ;:|||
     uProCysLeuSerCysSerSerValGlyAlaGlnHisLeuValSers 570

2260 ...............CTGGGTAATCCA............ 2249
 570 :::||||:::|||
     erThrGlyAlaSerGlyAspValGlySerProGlyPheGlyLeuProGly 586

2249 ...................................................2249

 587 LeuProGlyArgAlaGlyLyvalProGlyLeuLysGlyGluLysGlyAsnPh 603

2248 ..........CCTGCAGAGTCCCG...CATTCCAGTGCATGGAGCC 2215
 603 |||||  ||||||    |||  |||
     eGlyGluAlaGlyLyProAlaGlySerProGlyProProGlyProvalGlyP 620

2214 .......................................CCTGTA 2198
 620 ||:::||:::|||
     roAlaGlyIyIleLysGlyAlaLysGlyGluProCysGluProCysProAla 636

2197 TAAGTCCAGACTGAAACCCCCTGAAGGCCTCCAGTCAGGCAGC...... 2153
 637 |||::::|||||:::|||:::
     ..............LeuSerAsnLeuGlnAsnGlyAspValAr 646

2152 ..........CCTAGAGACTGGGGAGGAGAGGAGGGACGCCCCAGC 2116
 646 |||    |||||
     gValValAlaLeuProGlyProSerGlyGlu...LysGlyGluProGlyP 662

2115 CCCCAGCTGCTGCAGCTACGCACCTCAGCAGCAGGGTGGCAGCAGAGAG 2066
 662 ||||:::        |||||    |||
     roProGlyPheGlyLeu.....ProGlyLysGlnGlyLysAlaGlyGlu 676

2065 CCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCC 2017
 677 |||:::::     :::|||:::|||:::||||||:::
     ArgGlyLeuLysGlyGlnLysLysGlyAsp.AlaGlyAsnProGlyAspPro 693

2016 ....ATGGGGCTAACAGGAGCGGGG.....AGTGGGACCCA...GTG 1981
 693 |||   ||||||||      |||  |||
     lyThrProGlyThrArgGlyLeuSerGlyGluProGlyLyVal 709

1980 AGGCAGGCCCCTCCACCCAATGTGCTGGAAGTTTCTACGCTGAGTATTT 1931
 710 ::: |||||    ||       ||   :::
     Gln.GlyProAlaGlyProLysGlyGlyLysLysGlyAspGlyCysThrAlaC 726

1930 GGCCAAGTCGCTCTTGTGCAGCACCATAGGCAGTGACAGACTGGCTG 1881
 726 ||||||  :::  |||    :::
     ysProSerLeuGlnGlyThrValThrAspMetAlaGlyArgProGlyGln 742

1880 CCAGACCGCGCTGGCGAGCAGACCATCATAGGCAGTGACAGACTGGCTG 1831
 743 |||| |||    ::: |||
     ProGlyProLysGlyGluProGlyGluValGlyLyAlaGlyProGlyLyLy 759

1830 AGCTGGACAATGGAGACCCCATAAACAGGAGATGGGGCCACCTGGACAGCAG 1781
 759 |||||| ::: :::||||||||||||
     sProGlyGlnProGlyLeuProGlyValGlnGlyLyGlnProGlyLeuLysG 776
```

```
1100 GGCGCAGGGTGCGGGGCATGCGGCAGCAGCACAGCTGGTGCAGCCGGGGGAAGC 1051
     |||              :::       |||              |||
1033 GluGluGlyProProGlyMetArgGlySerProGlyProGlyProIl 1049

1050 AGGGCGCCCAGGT..........TCCGGAAAGCCAAGCGGCCCGGCA 1013
     |||    ||||||                      |||:::||||::  ||||||
1049 eGlyProProGlyPheProGlyAlaValGlySerProGlyLeuProGlyL 1066

1012 TGGACAGCAGTGGGGCGACAAGAGGGGCGCACAGCCCTTCTGCTGGCT 963
     ::::::||  |||||||  |||:::|||
1066 euGlnGly..........GluArgGlyLeuThrGlyLeuThrGlyAsp 1078

962 CGGTGGGGCCAGGCTGCCTCCTCAGCCA.................. 933
    :::   ||:::|| |||||
1079 LysGlyGluProGlyProGlyGlnProGlyTyrProGlyAlaThrGl 1095

932 ..........CCAGCAGTGCTGCTCACGCAGGTGAGGAAGATGAG 896
              |||::         ||     |||
1095 yProProGlyLeuProGlyIyIleLysGlyGluArg......GlyTyrThrG 1110

895 GGTGAGCAGGCCAAAGAGGCACTCCT..................... 870
    ||  |||||||:::||  |||
1110 lySerAlaGlyGluLysGlyGluProGlyProProGlySerGluGlyLeu 1126

869 CCTGGGTGCCCAGGT..........AGGGGGCCAGGGCA 841
    |||||  |||||                 ||||||||
1127 ProGlyProProGlyProAlaGlyProArgGlyGlyArgGlyPro..... 1141

840 CTGGTGTCCCAGTCAATGGCAGGCAGGAGTAGCCAGGCAGCCCCCAG 791
    |||              |||:::  |
1142 ..................GlnGlyAsnSerGlyGluGlyLysGlyA 1150

790 ACTGATCATGATGAAGGCATAGACAGAGTAGGCCTGGCGACAGTGGTCCGGGT 741
    ||                    |||||||       :::||||||
1150 sp..........GlnGlyPheGlnGlyGlnProGlyGlnProGlyProThr 1163

740 ..........CCCGGAAGAGGTCAGAGAGCAGGGCCTCCA..... 711
                   ||||||||
1164 GlySerProGlyPheProGlyLysValGlySerProGlyProProGlyPr 1180

710 ...GTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGCC 665
    ::||  :::||  |||  |||:::|||||::::::
1180 oGlnAlaGlyLysGlySerGluGluGlyIleArgGlyProSerGly...LeuP 1196

664 CAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATCCGGGCACAGCAGCC 615
    |||||           |||:::|||              ::::::
1196 roGlySerProGlyProPro.GlyProProGlyIleGlnGly........ 1209

614 CTGCTAGCCAGCCGGC..........CCTTGGGATGAGAAAGAGGCTCAGCAGG 571
    |||||||                 |||    |||::::::::
1210 ........ProAlaGlyLeuAspGlyLeuAspGlyLysAspGlyLysGlyProG 1224

570 ATGCCCAAGGACGAGTGCCCAGATGAAGGGCCGGCGGCG.....GCCATA 527
    |||  ::||| |||  |||||||||::      :::::::
1224 lyLeuArgGly..........AspProGlyProAlaGlyProProGlyLeu 1237

526 GGCGTCCACGCCAGTGGTCACTGGCTGCTGAGCCTAGGAGCCGGGACACAGACCA 477
    |||||              |||  |||
1238 MetGlyProProGlyPheProGlyGly..LysThrGlyHisProGlyLeuPro 1253

476 GGGCCAGCACTGGACCAATGCCCAGCCACCACATGGTCATGAACTTCTCCTCT 427
    |||||||       |||  |||        |||
1254 GlyProLysGlyLysAsp...CysGlyLysProGly............ 1263

426 ACCCCCA..........CTTCCAGCAGCAGGGCG............ 402
    |||||                |||::|||||:::
1264 .ProProGlySerThrGlyArgProGlyAlaGlyLeuGlyGluProGlyAlaM 1280

401 ..................GCACATAGGTGATGCCTGCGG 381
                      :::||        ||
1280 etGlyProGlnGlyArgProGlyProProGlyHis........ValGly 1293

380 CCAAACACACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCAAGAGCTGG 331
    |||
```

```
1294 ProProGlyProProGlyGln.................. 
     |||                    |||||||||||
     .........................ProGlyProAlaGl 1305

330 GCTTTCCGGT.....................................
    :::::||:::                                     321

1305 yIleSerAlaValGlyLeuLysGlyAspArgGlyAlaThrGlyGluArgG 1322

320 ......GCCCAGCAGGCGGCTCACCCACCAGCCCTGGACGCCATAGTGGG. 278
       |||:::                          |||  ||  ||
1322 lyLeuAlaGlyLeuProGlyGlnProGlyProGlyProGly.HisProGlyPr 1338

277 ..........CCAGGCGGGTAGGGCTGCAGGGGCCGTTCAGGCACTCCAG 238
            |||||                ::::::||:::::::||
1338 oProGlyGluProGlyThrAspGlyAlaAlaGlyAlaLysGlyLysGlyProProG 1355

237 AACTGCTTCGTCTCGCTCTGCTCCTCCAGAGAGCTGCGGCCCTCTCCTCCTTGC 188
    :::                          |||          :::
1355 lyLysGlnGlyPheTyrGlyLyProProGlyProLysGlyAspProGlyAla 1371

187 TGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTCTGCCAGCCCTT 138
    :::  |||||                |||            ||
1372 AlaGlyGln..........LysGlyGlnAlaGlyLysGlyGluLysGlyGlyArgAlaGlyMe 1386

137 TGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCTGTCTGTGGGG 88
    :  :::||||||:::|||::::::  |||  ::::::||:||||
1386 tProGlyGlyProGlyLysSerGlySerMetGly.ProValGlyProPro 1402

87 CACCTCAGTGGGGACACGTCTCATCATCAGATCCTGGCCAGGCGGCGCGCG 38
   :::||         |||:::::::  ::: |||  |||   ||||
1403 GlyProAlaGlyGlyHisProGlyHisProGlyAlaProGlyProSerGly.. 1418

37 GCTGTCACCCGGA 25
   ||||||||
1419 ....SerProGly 1421
```

seq_name: SwissProt_37:CA12_MOUSE

```
seq_documentation_block:
ID   CA12_MOUSE     STANDARD;     PRT;  1459 AA.
AC   P28481;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN   COL2A1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91358489.
RA   METSARANTA M., TOMAN D., DE CROMBRUGGHE B., VUORIO E.;
RT   "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT   structure, and alternative splicing.";
RL   J. BIOL. CHEM. 266:16862-16869(1991).
RN   [2]
RP   SEQUENCE OF 1455-1459 FROM N.A.
RX   MEDLINE; 91274355.
RA   METSARANTA M., TOMAN D., DE CROMBRUGGHE B., VUORIO E.;
RT   "Specific hybridization probes for mouse type I, II, III and IX
RT   collagen mRNAs.";
RL   BIOCHIM. BIOPHYS. ACTA 1089:241-243(1991).
CC   -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC   -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ----------------------------------------------------------------------
DR    EMBL; M65161; G200215; -.
DR    EMBL; X57982; E31338; -.
DR    MGD; MGI:88452; COL2A1.
DR    PROSITE; PS01208; VWFC; 1.
DR    PFAM; PF00093; vwc; 1.
KW    EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW    GLYCOPROTEIN; COLLAGEN; CARTILAGE; SIGNAL; ALTERNATIVE SPLICING.
FT    SIGNAL        1     25       POTENTIAL.
FT    PROPEP       26    153       AMINO-TERMINAL PROPEPTIDE
FT                                 (BY SIMILARITY).
FT    CHAIN       154   1213       COLLAGEN ALPHA 1(II) CHAIN.
FT    PROPEP     1214   1459       CARBOXYL-TERMINAL PROPEPTIDE
FT                                 (CHONDROCALCIN).
FT    DOMAIN       32     89       VWFC.
FT    DOMAIN      173   1186       TRIPLE-HELICAL REGION.
FT    DOMAIN     1187   1213       NONHELICAL REGION (C-TERMINAL).
FT    VARSPLIC     29     29       Q -> R (IN SHORT FORM).
FT    VARSPLIC     30     98       MISSING (IN SHORT FORM).
SQ    SEQUENCE   1459 AA;  139154 MW;  21053071 CRC32;

alignment_scores:
            Quality:  326.00          Length:    1101
              Ratio:   0.712            Gaps:      69
Percent Similarity:  41.599   Percent Identity:  26.612

alignment_block:
US-09-030-606-110/rev x CA12_MOUSE

Algn seg 1/1  to: CA12_MOUSE  from: 1  to: 1459

2557 GGCCACATCCTGATAAAAGGTAAGAGGGGGGTGGATCAGCAAAAGACAG 2508
        |||    ||  :::|||:::|||                |||||::
  32 GlySerCysLeuGlnAsnGlyGlnArg..........TyrLysAspLy  44

2507 TGCTGTGGGCTGAGGGACCTGGTTCTTGTGTGT...TGCCCTCAGGAC 2461
        :||  |||||||||||||     |||  :|||
  44 sAspValTrpLysProSerSerCysArgIleCysValCysAspThrGlyA  61

2460 TCTTCCCTACAAATAAGTCATATGTTCAAATCCATGGAGGAGTGTTTC 2411
        ::|||||||||      :::|||||     ::||||::|||
  61 snValLeuCysAspAspIleIleCysGluAspPro.....AspCysLeu  75

2410 ATCCTAGAAACTCCCATGCAAGAG.................. 2387
        |||  |||::|||   |||
  76 AsnProGlyIleProPheGlyCysCysProIleCysCysProAlaAspPle  92

2386 ......CTACATTAAACGAAGCTGCAGGTTAAGGGGCTTAGAGATGGGA 2344
             |||||||     ||||||| :::  ||:|
  92 uAlaThrAlaSerGlyArgLysLeuGlyProLysGly..GlnLysGlyG 108

2343 AACCAGGTGACTGAGTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTG 2294
        ::||||||||       ::||||||      :::|||
 108 luProGlyAspIleArgAspGlyAlaPro..GlyProArgGlyArgAsp 123

2293 TCTCAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGTAATCCA...CC 2247
        :::::|||||     ||       ||:||||||||| ||
 124 GlyGluProGly...................ThrProGlyAsnProGlyPr 134

2246 TGCAGAGTCCCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTAT 2197
        |||  |||      :|||| |||  |||  |||
 134 oAlaGlyProPro..............GlyProProGlyPro...... 144

2196 AAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGCAGC...... 2153
        ::||||||:::|:::|||::||:::
 145 ...................GlyLeuSerAlaGlyAsnPheAla 152

2152 ......................CCTAGAGACTGGGAGAGAGGAGAG 2127
                              ||  :::  ||||
 153 AlaGlnMetAlaGlyGlyTyrAspGlnIleGlyAlaGlyAlaGlnMetGl 169
```

```
2126 GACG..................................CCCCAGCGCCC 2113
        |:::                                 |||
 169 yValMetGlnGlyProMetGlyProMetGlyProArgGlyProProGlyP 186

2112 CAGCTGTGCAGCTACCGCACCTCAGCACCAGGCTGCAGCAGAGAGCCA 2063
        |||||                  :::::  ||||||||||||
 186 roAlaGlyAlaProGlyProGlyProGlyPheGlnGlyAsnProGlyGluPro 202

2062 CATTACTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCCATGG 2013
        |||       |||              |||       |||
 203 GlyGluProGlyValSerGlyProMetGlyProArgGlyProProGlyPr 219

2012 GGCTAACAGGAGCGGGGAGCTGGGA............CCCAGTGAGGCAG 1975
        ||:::::  |||:::        |||     ||:::::::::|
 219 oAlaGlyLysProGlyAspSerGlyGluAlaGlyLysProGlyLysSerG 236

1974 GC...........CCTCCACCCAATGTGTCTGGAAGTTTTCTACCCTGA 1937
        |||           :::::  ||||:::  |||::||||||
 236 lyGluArgGlyLeuProGlyProMetGlyAlaAArgGlyPhePro...... 250

1936 GTATTTGGCCAAGCTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAA 1887
        ||         :::||                 ||:::::
 251 .......GlyThrProGlyLeu..................ProGlyValLy 259

1886 TGGCGACCAGAC...CCAGGCTG..........CGGCAGA 1859
        :|||  :::     ||||||||||
 259 sGlyHisArgGlyTyrProGlyLeuAspGlyAlaLysGlyGluAlaGlyA 276

1858 CACCATATAGGCAGTGACAGACTGGCTGAGCTGGACAATGAGCCCATAA 1809
        ::||      :::  |||:::||  :::     :::
 276 laProGlyValLysGlyGluSerGlySerProGlyGluAsnGlySerPro 292

1808 ACAGGGATGGGGCCA........CCTGGACAGCAGGAAGGCACTATCC 1768
        ||||||          ||||| |||    |||
 293 GlyProMetGlyProArgGlyLeuProGlyGluArgGlyAlaArgThrGlyPr 309

1767 A..........GGATGGCGAGGTCCAGGCAGATGCCCGGCCGGAA 1731
             |||  :::      ||||||||  ||||||||
 309 oAlaGlyAlaAlaGlyAlaArgGlyAsnAspGlyInProGlyProAlaG 326

1730 ..CCACCCTGGCCTCG.......GTGGGCTCACCCACCACCACGTACG 1690
        |||||||                :::|||:::|||
 326 lyProProGlyProValGlyProAlaGlyGlyProGlyPheProGlyAla 342

1689 GAGACATCACAGGCAGGAGGCC.................CCGCAGAG 1661
        |||||||               |||:::
 343 ProGlyAlaLysGlyGluAlaGlyLysProThrGlyAlaArgGlyProGluGl 359

1660 CGCGGGTGGAGGTGGGAGCAGGCCACTGCCTCCAGCACCCACGTGTCCAT 1611
        :|||  ||||||||||||  :::||:::||:::   |||
 359 yAlaGlnGlySerArgGlyGluProGlyAsnProGlySerProGlyProA 376

1610 TAGGG...AAGGGAGCTCCAGGCTTAGGG.........CCTGGCAGGAAGCTG 1570
        |||    ::|||  |||||         |||||
 376 laGlyAlaSerGlyAsnProGlyThrAspGlyIlePheProGlyAlaLys.Gl 392

1569 GTCATCAGGCTGTCCTCACTGCTGTCTAGCACCTCCAGTGTCCCCTCGGTATTT 1520
        |||||:::|:::|||    ||||||      ||||||:::
 392 ySerAlaGlyAlaPro.....................GlyIle. 399

1519 GGGCAGGAACACCTGCTTCTCCCGGTGGTAGAGGGAGGCCAGTGTGTAGG 1470
        :|||||::|:::|||:::   |||||||     ||:::
 400 ..AlaGlyAlaProGlyPheProGlyProArgGlyProProGlyProGln 415

1469 GCA...................GGATCTGCAGGGCTGAGAAGGTGAACCCGGTG 1435
        :::                   |||  :::||||||||||||||||
 416 GlyAlaThrGlyProLeuGlyPro.LysGlyGlnAlaGlyGluProGlyI 432

1434 AGGGGCGGCTGAGCTGTCACCACGGC.....CACACTGTGGGACAGGCA 1391
        :::||  ||||||                 ::|||   |||
 432 leAlaGlyPheLysGlyAspGlnGlyProLysGlyGluThrGlyProAla 448
```

```
1390 TGTGGCACCGGCCAGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTG 1341
      ||| ||| ||| |||     ::|||
 449 GlyProGlnGly.AlaProGlyProAlaGlyGluGluGlyLysArgGlyAla 465

1340 CTCGA........GTGCCGAATCGCTGCACCACGGTCCATGACCAGAGAG 1297
     |||||         ::: |||::||| |||||| |||||
 465 laArgGlyGluProGlyGly.AlaGlyProIleGlyProProGlyGluAr 481

1296 AAGA............CCAGGGAGATGGCGCACTGCAGGAA 1268
     |||||||         ::||||
 481 gGlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGly. 497

1267 CAGCCCCAGGTGCCCATCCGAACGCCTTCATCATAGTGTCTCCGGCCT 1218
     |||||::||:: |||     ::|||
 498 ..ProLysGlyAlaProGlyGluArgGlyLeuProSerGlyLeuAlaGlyPro 513

1217 CGG...........TGCCCGGCTCAG........... 1203
     :::|
 514 LysGlyAlaAsnGlyAspProGlyLyArgProGlyLeuProGlyLeuProGl 530

1202 ............CTCTGGGCACGCCCTGGTACAGCC............ 1179
     |||  |||||||:::|||
 530 yAlaArgGlyLeuThrGlyLyArgProGlyAspAlaGlyLyProGlnGlyLysv 547

1178 ...................CCTCGCCCACGAAAT 1164
                        ||| ||| |||
 547 alGlyProSerGlyAlaProGlyLyGluAspGlyLyArgProGlyProGly 563

1163 CCGTGTAAAAACAGCGTGAAGGTCATGAGTG............CCATC 1129
     |||                                         |||::
 564 ProGlnGlyAlaArgGlyGlnnProGlyValMetGlyPheProGlyProLy 580

1128 CAGTCGCACAGCTCAGCCACGAAGACCGGC...........GCAGGGT 1091
     ::|:::|||:::|||  :||||||                  |||||||
 580 sGlyAlaAsnGlyGluProGlyLyLysAlaGlyGluLysGlyLeuAlaGlyA 597

1090 GCGGGGCATGCGGCGACCAGCTGGTGCAGCCCGGGAAGCCAGGGCGCCCA 1041
     :: :::              |||  |||||||:::|||
 597 laProGlyLeuArgGlyLeuProGlyLyLysAspGlyGluThrGlyAlaAla 613

1040 GGTTCCGGAAAGCCAAGCGGGCCCGGCCATGGACCAGCAGCAGCGGCGACAAG 991
     |||  ||||||||||||:::               ::::
 614 GlyProProGlyProSerGlyYProAlaGlyGluArgGlyGlyGlnGlyAl 630

 990 GAGGGGGCCCGACAGCCCTTCTGCTGGTCGGTGGGGCCGGCCTGCCTC 941
     |||||||||                  |||
 630 aProGlyProSerGlyPhe..........GlnGlyLeuProGlyProP 643

 940 CTCAGCCACCA............ 930
     || ||||||
 643 roGlyProProGlyGluGlyLyLysGlnGlyAspGlnGlyLeuGlyProGlyL 659

 869 ..CCTGGGTGCCCAGGT..........AGGGGG........... 849
     |||||| |||||                ::|||
 869 euProGlyThrProGlyThrAspGlyProLysGlyAlaAlaGlyProGlyAsp 709

 848 ......CCAGGGCACTGGTGTCCCAGTCAATGGCAGGCA............ 816
     |||||| |||||                |||:::
 848 GlyProProGlyLyAlaGlnGlyProProGlyLeuGlnGlyMetProGlyGl 726

 815 ...................GGAGGTAGCCAGGCAGCCCCCAAGACTGAT 785
                        :::|||  ||| ||
 726 uArgGlyAlaAlaAlaGlyIleAlaGlyLyProLysGlyAspAlaGlyValG 743

 784 CATGAAGGCCATAGACAGAGTAGGCCTGGCGACAGTGGT............ 747
     |
```

```
 743 lyGluLysGlyProGluGlyProGlyLyAspGlyLyLysGlyArgGlyArggGlyLeu 759
     ||||| :::|||
 746 .........................CCGGGTCCC..........GGAAGAGGTCA 727
                             |||||||||
 760 ThrGlyProIleGlyProProGlyProAlaGlyAlaAsnGlyGluLuLysGl 776

 726 GAGAGCAGGGCCTCCA...GTGGAGTGAAGCACA............ 696
     ::|||||||||         |||:::
 776 yGluAlaGlyProProGlyProSerGlySerThrGlyAlaArgGlyAlaP 793

 695 ........CCTGGGCCACAGAAGTCCAGCAGCCCCAGGC. 666
     |||||: ::||| ::||||
 793 roGlyGluProGlyGluThrGlyProProGlyProAlaGlyPheAlaGly 809

 665 ...CCAGGATGAGCAGTGCCAGCTCCAGGGGCC....TGGGATCCGGGCAC 622
     ||||| |||||      |||||||
 810 ProProGly..AlaAspGlyGlnProGlyAlaLysGlyAspGlnGlyGl 825

 621 AGCAGCCCTGCTAGCCAGCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAG 572
     |||:::            :::|||:::|||:::|||
 825 uAlaGly.....................GlnLysGlyAspAlaG 833

 571 GATGCCCAAGG............... 561
     ||::::|||
 833 lyAlaProGlyLyProGlnGlyProSerGlyAlaProGlyLyProGlnGlyPro 849

 560 .........ACAGTGCCCAGATGAGAAGGGCCGGCCGGCCCATAGCGTCCA 520
     |||  ||||||:::
 850 ThrGlyValThrGlyLyProLysGlyAlaArgGlyValArgGlyProProGl 866

 519 CGGCAGTGTCACTGGCTGGCTGAGCCTAGGACGGGACACCAGGACCAGGC...C 473
     ||||:::||        |||||:::|||  ||| |||
 866 yAlaThrGlyPhe.......ProGlyAlaAlaGlyLyArgValGlyProP 880

 472 CAGCACTGGACCAATGCCCAGCACCATGGTCATGAACTTCTCCTCTACCC 423
     ||:::         |||::||| :::
 880 roGlyAlaAsnGlyAsnProGlyProAlaGly.............Pro 891

 422 CCACTTCCAGCAGAGGCGGCACATAGGTGATGCCTGCGGCCAAACAC 373
     ||| |||||||:::                 :::
 892 ProGlyProAlaGlyLyAspGlyLyProLysGlyValArgGlyAspSerGl 908

 372 ACCTCCAGGCCAAAGGTTAGCAGGTTGACCAGCAAGAGCTGGCTTTCCG 323
     |||||||||              ||||||| |||  :::|||
 908 yProProGlyArg.......AlaGlyAspProGlyLeuGluGlyProAlaG 923

 322 GTGCC...GCAGCAGGCGGCTCACCCACCAGCCTCTGGACCATAGTGGG.. 278
     |||| :::||:::|||       :::|||||| |||||
 923 lyAlaProGlyGluLysGlyGluProGlyAspAspGlyPro.SerGlyLe 939

 277 .........CCAGGCGGTAGGGCTCAGGGCCCGTTCAGGCACT.... 242
     |||                ::::
 939 uAspGlyProProGlyProGlnGlyLeuAlaGlyGlnArgGlyIleValG 956

 241 ...............................CCAGAACTGCTTCGTCTC 224
                                    ||| |||
 956 lyLeuProGlyGlnArgGlyLeuGlnGlyPheProGlyLeuProGlyPro 972

 223 GGCTCTGCTGCTCCAGGAAGCTGCGGCCCTCCTCCCTTGCTGCCCAACTGCC 174
     |||||||||||||        :::::::||||
 973 .SerGlyGluProGlyLysGlnGlyLyAlaProGlyAlaSerGlyAsp..... 987

 173 TAGGAATCAGCCAGGGCGCCCATTTCTGCCAGCCCTTTGGTGCCGGTCCAG 124
     |||::: |||||                 |||
 988 .ArgGlyProProGly....................ProValGly.ProProGly 998

 123 CTTCTCAGCCCCATGCTCAAACCTGCTGCTGTGGGCACCTCAGTGGGA 74
     ||| ::::||:::|||::::    ||| ||| :::::::||
 999 LeuThrGlyProAlaGlyLyProLysGlyArgGluProGlyAlaAs 1015

  73 CACGTCTCATCACTCAGATCCTGGCCGA............ 46
     ::: ||||||||
```

```
1015 p..............GlyProProGlyArgAspGlyAlaAlaGlyValLysG 1028
                        ::||||||
  45 ................GGCCGCGGCTGTCACCCGAGCC 22
                        :::||||||||
1028 lyAspArgGlyGluThrGlyAlaLeuGlyAlaProGlyAla 1041
```

seq_name: SwissProt_37:STP_SPIOL

seq_documentation_block:
```
ID  STP_SPIOL   STANDARD;       PRT;   525 AA.
AC  Q03411;
DT  01-JUN-1994 (REL. 29, CREATED)
DT  01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  SUCROSE TRANSPORT PROTEIN (SUCROSE PERMEASE) (SUCROSE-PROTON
DE  SYMPORTER).
OS  SPINACIA OLERACEA (SPINACH).
OC  EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC  EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC  CARYOPHYLLIDAE; CARYOPHYLLALES; CHENOPODIACEAE; SPINACIA.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=LEAF;
RX  MEDLINE; 93099843.
RA  RIESMEIER J.W., WILLMITZER L., FROMMER W.B.;
RT  "Isolation and characterization of a sucrose carrier cDNA from
RT  spinach by functional expression in yeast.";
RL  EMBO J. 11:4705-4713(1992)
CC  -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE
CC      CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).
CC      CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.
CC  -!- PATHWAY: SUCROSE METABOLIC PATHWAY.
CC  -!- SUBCELLULAR LOCATION: INNER MEMBRANE.
CC  -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR  EMBL; X67125; G21319; -.
DR  PIR; S28052; S28052.
DR  PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR  PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
DR  PFAM; PF00083; sugar_tr; 1.
KW  DUPLICATION; TRANSMEMBRANE; TRANSPORT; SUGAR TRANSPORT; SYMPORT;
KW  GLYCOPROTEIN.
FT  DOMAIN        1     40       CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM     41     60       POTENTIAL.
FT  TRANSMEM     70     88       POTENTIAL.
FT  TRANSMEM    108    127       POTENTIAL.
FT  TRANSMEM    146    165       POTENTIAL.
FT  TRANSMEM    185    204       POTENTIAL.
FT  TRANSMEM    230    249       POTENTIAL.
FT  TRANSMEM    296    315       POTENTIAL.
FT  TRANSMEM    341    360       POTENTIAL.
FT  TRANSMEM    373    392       POTENTIAL.
FT  TRANSMEM    419    438       POTENTIAL.
FT  TRANSMEM    441    461       POTENTIAL.
FT  TRANSMEM    490    509       POTENTIAL.
FT  DOMAIN      510    525       CYTOPLASMIC (POTENTIAL).
FT  CARBOHYD    270    270       POTENTIAL.
FT  CARBOHYD    271    271       POTENTIAL.
SQ  SEQUENCE    525 AA;   54992 MW;   000477CA CRC32;
```

alignment_scores:
```
        Quality:  327.00        Length:     571
          Ratio:    1.168         Gaps:      22
Percent Similarity:  49.037    Percent Identity:  23.468
```

alignment_block:
US-09-030-606-110 x STP_SPIOL

Align seg 1/1 to: STP_SPIOL from: 1 to: 525

```
310 CCTGCTGGGCACCGGAAAGCCCAGCTCTTGCTGGTCAACCTGCTAACCT 359
    ||:|||  |||||| ||:::  |||||  ||:::
 25 ProThrThrProProGlu.AlaGluAlaThrLeuLysLysLeu........ 38

360 TTGGCTGGAGGTGTTGGCCCAGGCATCACCTAT................ 397
    ||:||   ::: |||:||||:||||||  ::::
 39 ..GlyLeuValAlaSerValAlaAlaGlyValGlnPheGlyTrpAlaLeu 54

398 ..............GTGCCCCTCTGCTGTGGAAGTGGGGTAGAGGAGAA 435
                  |||  ||   |||:|||||:|||  :::
 55 GlnLeuSerLeuThrProTyrValGlnLeuLeuGlyIleProHisTh 71

436 GTTCATGACCATGTGCTGGACAGGTCCAGTGCTGGGCCTGGTCTGTG 485
    :::  ||   ||| ||:  ||  ||:: |||::  |||::||||||
 71 rTrpAlaAlaTyrIleTrpLeuCysGlyProIleSerGlyMetIleValG 88

486 TCCCGCCTCTAGGTCAGCCAGTGACCACTGGCGTGGAGCGTATGGCCGC 535
    ||||||:::||    ||||||   ||||||||||  :::|||::||||||
 88 lnProLeuValGlyTyrTyrSerAspArgCysThrSerArgPheGlyArg 104

536 CGCCGGCCCTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTT 585
    |||||||||||||| ||||  |||||||||||:::|||||:::|||::|||
105 ArgArgProPheIleAlaLaGlyAlaAlaLeuValAlaValAlaValGl 121

586 TCTCATCCCAAGGGCCGGCTGGCTAGCA.........GGGCTGCTGCCCGG 629
    |||||| ||| |||||||||||  |||       |||:||||||:||||
121 yLeuIle.........GlyPheAlaAlaAspIleGlyAlaAlaSerGlyA 135

630 ATCCC................AGGCCCCTGGAGCTGGCACTGCTCATCCTG 664
    |||||               ::||||:||||||||:::::::::::::::
135 spProThrGlyAsnValAlaLysProArgAlaIleAlaValPheValVal 151

665 GGGCTGGGGCTGCTGGACTTCTGTGGCCAGGTGCTTCACTCCACTGGA 714
    || ||||||||||||  ||||||||||||||||  ||||  ::
152 GlyPheTrpIleLeuAspValAlaAsnAsnThrLeuGlnGlyProCysAr 168

715 GGGCCTGCTCTGACCTCTTCCCG...GACCCCGACCCACTGTCGCCAGG 761
    :||||||||||||| |::   ||   :::   |||||||||:||||||:
168 gAlaLeuLeuAlaAspMetAlaAlaGlySerGlnThrLysThrArgTyrA 185

762 CCTACTCTGCTCTATGCCTCATGATCAGTCTTGGGGCTGCCTGGCTAC 811
    ||  :::  :::|||::||:|||   ||||||||||||||:: ||||||
185 laAsnAlaPheSerPhePheMetAlaLeuGlyAsnIleGlyGlyTyr 201

812 CTCCTGCCTGCCATTGAC.................TGGGACAC 837
    |||||||||||||  ||:                 ||
202 AlaAlaAlaGlySerTyrSerArgLeuTyrThrValPheProPheThrLysTh 218

838 CAGTGCCCTGGCCCCTACCTGGGCACCCGAGGAGTGCCTCTTGGCC 887
    |:::||   ::: |||| :::  ||::|||||||||:::::|||||
218 rAlaAlaCysAspValTyrCysAlaAsnLeuLysSerCysPheIleS 235

888 TGCTCACC...CTCATCTTCCTCACCGCGTAGCAGCCACACTGCTGGTG 934
    ::::|||   ||||||| |||||| |||:: :::  :::::||||||||
235 erIleThrLeuLeuIleValLeuThrIleLeuAlaLeuServalValLys 251

935 GCTGAGGAGGCAGCGCGTGGGCCCCACCGAGCCAGCAGAAGGGCTGTCGGC 984
    ||||:::  ||||||  |||  ||  ||| :::  |||  |||||||||||
252 GluArgGlnIleThrIleAspGluIleGluGlyLeuAspLeuLysAs 268

985 CCCCTCCTTGTCGCCCATGCCGTGCCATGCCGGGCCGCTTGGCTTTCC 1034
    ::::      |||       ||:: :::  ||||||||  |||||||||
268 nArgAsnAsnSerGlyCys.........AlaArgLeuProPhe. 280

1035 GGAACCTGGCGCCCTGCTTCCCGGCTGCACCAGCTGTGCTGCCCGCATG 1084
     ::::|||  |||  |||||||  ||||:|| ::||||
281 .....PheGlnLeuIleGlyAlaLeuLysAspLeu............. 291
```

seq_name: SwissProt_37:CA25_HUMAN

```
seq_documentation_block:
ID   CA25_HUMAN     STANDARD;      PRT;   1496 AA.
AC   P05997;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   PROCOLLAGEN ALPHA 2(V) CHAIN PRECURSOR.
GN   COL5A2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE OF 1-463 FROM N.A.
RX   MEDLINE; 89123368.
RA   WOODBURY D., BENSON-CHANDA V., RAMIREZ F.;
RT   "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT   to the structural criteria of a fibrillar procollagen molecule.";
RL   J. BIOL. CHEM. 264:2735-2738(1989).
RN   [2]
RP   SEQUENCE OF 398-1496 FROM N.A.
RX   MEDLINE; 87146331.
RA   WEIL D., BERNARD M.P., GARGANO S., RAMIREZ F.;
RT   "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT   major fibrillar-forming collagens.";
RL   NUCLEIC ACIDS RES. 15:181-198(1987).
RN   [3]
RP   SEQUENCE OF 1227-1496 FROM N.A.
RX   MEDLINE; 85289937.
RA   MYERS J.C., LOIDL H.R., SEYER J.M., DION A.S.;
RT   "Complete primary structure of the human alpha 2 type V procollagen
RT   COOH-terminal propeptide.";
RL   J. BIOL. CHEM. 260:11216-11222(1985).
RN   [4]
RP   SEQUENCE OF 1449-1496 FROM N.A.
RX   MEDLINE; 89138450.
RA   TSIPOURAS P., SCHWARTZ R.C., LIDDELL A.C., SALKELD C.S., WEIL D.,
RA   RAMIREZ F.;
RT   "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT   located on the long arm of human chromosome 2.";
RL   GENOMICS 3:275-277(1988).
RN   [5]
RP   SEQUENCE OF 208-227.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 92239022.
RA   MANN K.;
RT   "Isolation of the alpha 3-chain of human type V collagen and
RT   characterization by partial sequencing.";
RL   BIOL. CHEM. HOPPE-SEYLER 373:69-75(1992).
RN   [6]
RP   SEQUENCE OF 288-297 AND 606-617.
RC   TISSUE=BONE;
RX   MEDLINE; 94237164.
RA   MORADI-AMELI M., ROUSSEAU J.C., KLEMAN J.P., CHAMPLIAUD M.F.,
RA   BOUTILLON M.M., BERNILLON J., WALLACH J.M., VAN DER REST M.;
RT   "Diversity in the processing events at the N-terminus of type-V
RT   collagen.";
RL   EUR. J. BIOCHEM. 221:987-995(1994).
CC   -!- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE
CC       COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS
CC       TO DNA, HEPARAN SULFATE, THROMBOSPONDIN, HEPARIN, AND INSULIN.
CC   -!- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN
CC       MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND
CC       ONE ALPHA 3(V) CHAINS IN PLACENTA.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on  its
```

```
1085 CCCCGCACCCTGCGCCCGGCTCTTCGTGCTGAGCTGTGCAGCTGGATGGC 1134
     |||::: ::  |||    :::||::|:|
292  ProLysProMetLeuIleLeuLeuIleLeuValThrAlaLeuAsnTrpIleAl 308

1135 ACTCATGACCTTCACGCGTGTTTACACGGATTTCGTGGGCCAGGGGCTGT 1184
     :::  ||| ||| |||:::::|||| :::|
308  aTrpPheProPheLeuLeuPheAspThrAspTrpMetGlyLysGluValT 325

1185 ACCAGGGCGTGCCCAGAGTCGACCGGCACCGGGCACCCGGAGACACTAT 1234
     ||  |||::        :::  |||
325  yrGlyGlyThr..........ValGlyGluGlyLysLysLeuTyr 335

1235 GATGAAGGCGGTTCGGATGGGCAGCCTGGGCTGTTCCTGCAGTGCGCCAT 1284
     ||::||||||:: :|||||||||||| :::::  :
336  AspGlnGlyValHisAlaGlyLeuGlyLeuMetIleAsnSerValVa 352

1285 CTCCCTGGTCTTCTCTCGTCATGACCGGCTGTGCAGCGATTCGGC 1333
     |||:::|||::  :::::||||| ||||||| |
352  lLeuGlyValMetSerLeuSerIleGluGlyLeuAlaArgMetValGlyG 369

1334 ....ACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTTCCCTGTGGCT 1378
     |||
369  lyAlaLysArgLeuTrpGlyIleValAsnIleIleLeu............ 381

1379 GCCGGTGCCCATGCCGTGTCCCACAGTGGCCGGTGCAGCTTCAGC 1428
     |||:::|||||::  :::||| |||||||||| ||||
382  ....AlaValCysLeuAlaMetThrVal...LeuValThrLysSerAl 395

1429 CGCCCCTCACCGGGTTCACCTTCTCAGCCCTGCAGATCCTGCCCTACACAC 1478
     ||
395  aGlu.......... 396

1479 TGGCCCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATACCGA 1528
     :: |||:::::::  ::::::
397  ...HisPheArgAspSerHisHisIleMet.......... 405

1529 GGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCT 1578

405  .......... 405

1579 GCCAGGGCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGTGCTG 1628

405  .......... 405

1629 GAGGCAGTGGCCTGCTCCCACCTCCACCGCGCTCGCGGGCCCTCTGCC 1678
     |||||::::::||  ||||||||||||
406  ..GlySerAlaValProProProProProAla.....GlyValLysGly 419

1679 TGTGATGTCTCCGTACGTGTGGTGGTGGTGGGCCACCCAGGCCAGGGT 1728
     ::||||||::  :::|||::::::|| |||
420  GlyAlaLeuAlaIlePheAlaValLeuGlyIleProLeuAlaIleThrPh 436

1729 G...GTTCCG.........................GGCCGGG 1742
     ||  ||:|||
436  eSerIleProPheAlaLeuAlaSerIlePheSerAlaSerGlySerG 453

1743 GCATCTGCCTGGACCTCGCGCCATCCTGGATAGTGCCTTCCTGCTCCCAG 1792
     ||  |||||:::::::::::::  |||||
453  lyGlnGlyLeuSerLeuGlyValLeuAsnLeuAlaIleValProGln 469

1793 GTG.......... 1795
     :::
470  MetPheValSerValThrSerGlyProTrpProAspProAlaMetPheGlyGlyGl 486

1796 ....GCCCCATCCTGTTTATGGCCTCCATTGTCCAGTCCAGCCAGTCTG 1841
     |||::::::: :::|||  |||
486  yAsnLeuProAlaPheValValGlyAlaValAlaLeuAlaThrAlaSerAlaV 503

1842 TCACTGCCTAT 1852
     ||  :::::
503  alLeuSerPhe 506
```

```
CC   use  by  non-profit  institutions  as  long  as  its  content  is  in  no  way
CC   modified  and  this  statement  is  not  removed.  Usage  by  and  for  commercial
CC   entities  requires  a  license  agreement  (See  http://www.isb-sib.ch/announce/
CC   or  send  an  email  to  license@isb-sib.ch).
DR   EMBL;  J04478;  G179698;  -.
DR   EMBL;  X04758;  E3803;  -.
DR   EMBL;  M11718;  G180913;  -.
DR   PIR;  A25374;  A25374.
DR   PIR;  A25874;  A25874.
DR   PIR;  A30017;  A30017.
DR   PIR;  A31427;  A31427.
DR   MIM;  120190;  -.
DR   PROSITE;  PS01208;  VWFC;  1.
DR   PFAM;  PF00093;  vwc;  1.
KW   EXTRACELLULAR  MATRIX;  CONNECTIVE  TISSUE;  REPEAT;  HYDROXYLATION;
KW   GLYCOPROTEIN;  COLLAGEN;  SIGNAL.
FT   SIGNAL        1      26
FT   CHAIN        27    1226       COLLAGEN  ALPHA  2(V)  CHAIN.
FT   PROPEP     1227    1496       CARBOXYL-TERMINAL  PROPEPTIDE.
FT   DOMAIN       39      97       VWFC.
FT   MOD_RES     290     290       HYDROXYLATION.
FT   MOD_RES     293     293       HYDROXYLATION.
FT   MOD_RES     296     296       HYDROXYLATION.
FT   MOD_RES     608     608       HYDROXYLATION.
FT   MOD_RES     614     614       HYDROXYLATION.
FT   CONFLICT    292     292       A  ->  P  (IN  REF.  6).
FT   CONFLICT   1418    1418       K  ->  T  (IN  REF.  3).
FT   CONFLICT   1438    1438       F  ->  S  (IN  REF.  3).
FT   CONFLICT   1460    1460       E  ->  Q  (IN  REF.  4).
FT   CONFLICT   1496    1496       V  ->  A  (IN  REF.  4).
SQ   SEQUENCE   1496  AA;  144720  MW;  482DAAEF  CRC32;

alignment_scores:
          Quality:   324.50          Length:  1196
            Ratio:     0.702           Gaps:    78
Percent  Similarity:  38.629       Percent  Identity:  25.418

alignment_block:
US-09-030-606-110/rev  x  CA25_HUMAN  ..

Align  seg  1/1  to:  CA25_HUMAN    from:  1   to:  1496

3068  GGTCCTGCAGTAGCTCCAAACAGGGTTGTGGACTGGTGGGGAAAGTTGG  3019
      ||||||  ||  |||::::|||:::                  |||  |||:::|
 216  GlyPro...ValGlyProArgGlyProGlnGlyLeuGlnGlyGlnGlnGl  231

3018  GGGTAGGGGAAAGTTGGG...GTAGGGGAAATTTGGCAGTGCCTTCA  2972
      ||||  |||  |||  |||||          ||:::  :        ::  |||
 231  yGlyAlaGlyProThrGlyGluProProGlyGluProGlyAspProGlyProM  248

2971  TCAGCCCAGTCCTAGAGAGA...GTAGAGGGG...............  2943
      ::::::|||::|||  ::|||     ||||||
 248  etGlyProIleGlySerArgGlyProGluGlyProGluGlyProLysProGly  264

2942  ...AGTGAAGTGGGGGAACCAGGCTGGGCCAAGAGAAGAGGGGGTGGTT  2896
      ::::::|||  |||:::::::::|||          |||  |||:::
 265  GluAspGlyGluProGlyArgAsnProGlyGluValGlyPheAl  281

2895  AGGGAAGCCGTTGAGA...........CCTGAAGCCCACCCTACCTTCC  2855
      |||:::|||              .................GlyLeuPro... 296
 281  aGlySerProGlyAlaArgGlyPheProGlyAlaPro.GlyLeuPro...  296

2854  TTCAACACCCTAACCTTGGGTAACAGCATTTGGAATTATCATTTGGGATG  2805
      |||||::|||:::::::|||          |||  |||
 297  ..............GlyLeuLysGlyHisArgGlyHisLysGly..  306

2804  AGTAGAAATTTCCAAGGTCCTGGGTTAGGCATTTTGGG..............  2768
      ::|||:|||:|||  |||  ||||||            |||
 307  .LeuGluGlyProLysGlyValGlyValGlyAlaProGlySerLysGlyGluA  323

2767                              GGGCCA...GACCCCAGGAGA  2750
```

Right column (continuation of alignment):

```
 323  laGlyProThrGlyProMetGlyAlaMetGlyProLeuGlyProArgGly  339
                                                     ||||||
2749  AGAAGATTCGGCAATGACATGATCAGCCCAATGACCAGCTATCCAGGGGACCT  2700
                              |||:::|         |||
 340  Met...........................ProGlyGluArgGlyArgLe  347

2699  GATTGTTGGGGATCCCCACCCTACCCAAATATTAGACACCAACACAGA  2651
      |  |||          |||:::||  |||        |||
 347  u...........GlyProGlnGlyAlaProGlyGlnArgGlyAlaHisGlyM  361

2650  ...............................AAAGCTAGCAAT  2639
                                      :::||||||
 361  etProGlyLysProGlyProMetGlyProLeuGlyIleProGlySerSer  377

2638  GGATCCCCTCTACTTGTTGTTAAATAAATAAGTTAAATATTAAATGCCTG  2589
      ||||||||:::::|
 378  GlyPheProGlyAsn..................................  382

2588  TGTCTCTGTGATGGCAACAGAGGACCAACAGGCCACATCCTGATAAAAG  2539
      :::::::||  :::  |||||||||||
 383  .ProGlyMetLysGlyGluAlaGlyProThrGly........AlaArgG  396

2538  GTAAGAGGGGGTGATCAGCAAAAGACAGTGCTGTGGGCTGAGGGGAC  2489
      |||:::::|||          |||  |||  ||||||||||
 396  lyProGluGlyProGlnGlyArgGlyThrGlyGluProProGlyPro  412

2488  CTGGTTCTGTGTGTGCCCTCAGGACTCTTCCCTACAAATAAGTCAT  2439
      :::|||::::::::
 413  ValGlySerProGlyLeuProGlyAlaIleGly.................  423

2438  ATGTTCAAATCCATGGAGGAGTGTTCATCCTAGAAACTCCATGCAAG  2389
 423  .................................................  423

2388  AGCTACACATTAAACAAGCTGCAGGTTAAGGGGCTTAGAGATGGGAAACCA  2339
                                         |||  |||
2338  GGTGACTGAGTTATTCAGCTCCCAAAAACCCTTCTTCAGGTGTGTCTCA  2289
      ||||||:::::::::|||:::
 424                                 .....ThrAspGlyThrPro  428

2338  GGTGACTGAGTTATTCAGCTCCCAAAAACCCTTCTTCAGGTGTGTCTCA  2289
      |||:::||||
 429  Gly.................................ProLysGlyProThr.  434

2288  ACTAGGAGGCTAGCTGTTAACCCTGGGTAATCCACCTGCAGAGT  2239
                                         |||:::||||||
 435  ........................GlySerProGlyThrSerG  441

2238  CCCCGCCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAG  2189
      :::|||          :::||||||          |||
 441  lyProProGlySerAla..GlyProProGlySerProGly...ProGln  455

2188  ACTGAAACCCCCTGGAAGGCCTCCAGCCTCAGGCAGCCCTAGAGACTGGGG  2139
 456  GlySerThrGlyProGlnGly...AsnSerGlyLeuProGlyAspProGl  471

2138  AGAGAGG...............AGAGGGACGCCCCCAGCCTGTGC  2104
             .::||              :::|||  |||  |||:::::||
 471  yPheLysGlyGluAlaGlyProLysGlyGlyProHisGlyIleG  488

2103  AGCTACGCACCTCAGCAGCACAGGTGGCAGCAGAGCCACATTACTTT  2054
      ||
 488  ln...............................................  488

2053  GGCAGCAACAGAAACTGGGCGGCAGCCGCAGCCGCCATGGGCTAACAG  2004
                |||          |||
 489  ..................GlyProIleGlyProProGlyGluGlyGlyLy  499

2003  GAGCGGGGAGCTGGGA...CCCAGTGAGGCGAGGGCCCTCCACCCAATGTG  1957
      :|||||||:|||  :::|||  |||:||||||||||||
 499  sArgGlyProArgGlyAspProGlyThrLeuGlyPro.........  512

1956  CTGGAAGTTTTCTACGCTGAGTATTTGGCCAAGTCGCTCTTGTCAAATAC  1907
                              |||      |||
```

```
513  ...................................GlyProVal...................  515

1906  TACCTGTGTAGCAAAGTAAATGCGACCAGACCCAGGCCTGCGGCAGACA  1857
         |||:::||                        ::|||
516   .............GlyGluArgGlyAlaProGlyAsnArgGl  525

1856  CCATATAGGCAGTGACAGACTG..............GCTGAGCTGGACA  1822
         |||||||||    |:::                ||||:::  :::
525   yPheProGlySerAspGlyLeuProGlyProLysGlyAlaGlnGlyGluA  542

1821  ATGGAGCCCATAAACAGGGATGGGCCACCTGGGACACGAGGAAGGCACT  1772
         :|||  |||:::::  :::  ||||| |||:::||     ||
542   rgGly..ProValGly.SerSerGlyProLysGlySerGlnGly......  555

1771  ATCCAGGATGGGCGAGGTCAGGCAGATGCCCGGCCCGGAACCACCCTGG  1722
         ||||||||          ||| :::  ||||| |||||
556   .............AspProGlyArgProGlyGluPro.GlyLeuProGl  567

1721  CCTCGGTGGCCTCACCCACCACCACGTACGGAGACATCACAGGCAGA.  1673
         ||||||||       ::     |||:::          ||
567   yAlaArgGlyLeuThrGlyAsnProGlyValGlnGlyProGluGlyLysL  584

1672  ..GGCCCCGAGAGCGCGGGTGGAGGTGCAGGCCACTGCCTCCAGC  1625
         |||||| |||    :::|||||  ::|||||| ||||||
584   euGlyProLeu.GlyAlaProGlyGluAspGlyAsnProGlyArgProProGl  600

1624  ACCCAGTGTCCATTAGGG....AAGGGAGCTCCAGGC...TTAGGG...C  1584
         :::|||             ::|||                  |
600   y.........SerIleGlyIleLysGlyGlnProGlyThrMetGlyLeuP  614

1583  CTGGCAGGAAGCTGGTCATCAGGCTGTCCTCCTCACTGCTAGCACCTCCAGTG  1534
         |||||||| |||    |||:::|| |||        ||
614   roGlyProLys.GlySerAsnGlyAspPro..................  623

1533  TCCCCTCGGTATT....TGGGCAGGAACACCTGCTTCTCCGGTGTA  1490
         |||            ||  |||||::|||:::||  |||||
624   .....GlyLysProGlyGluAlaGlyAlaGlyAsnProGlyValProGlyGlnA  638

1489  GAGGGAGG....................CCAGTGTGTAGGGCAGGA  1464
         ||||||                    |||
638   rgGlyAlaProGlyLysProGlyAspGlyLysValGlyProTyrGlyProProGly  654

1463  TCTGCAGGGCTGAGAAGGTGAACCCCGTGAGGGCGGCTGAGAGCTGTCACC  1414
         |||||||| ||||| |||||| |||||   |           ::|||
655   ProProGlyLeuArg.GlyGluArgGlyGluGlnGly..ProProGlyPr  670

1413  ACGGCCACACTGTGGGACAGGCATGTGGCACCGGCACGCCACCAGGGAAAG  1365
         |||                ||||||::::|||     :::||||||
670   oThrGlyPheGlnGly..........HisProGlyProProGlyProP  683

1364  ..............................CTGCCACACTGGCCAAAT  1347
         :::|||               |||:::
683   roGlyGluGlyGlyLysProGlyAlaGlyAspGlnGlyValGlyProGly  699

1346  AGACTGCTCGAGTGCCGAATCGCTGCACCAGCCGGTGCCATGACCAGAGAG  1297
         ::|||||  ||||| ||||||          ||||
700   AlaVal.............GlyProLeuGlyProArgGlyGluAr  710

1296  A...AGACCAGGAGAGATGCGCACTGCAGGAACAGCCCCAGGCTGCCCAT  1250
         ::|||||||:::                    :::  ||||
710   gGlyAsnProGlyGluArgGlyGluProGlyIleThr...GlyLeuProG  726

1249  CCGAACGC........CTTCATCATAGTGTCCGGGCCTCGG.......  1215
         ::|||:::        :::||||||:::::
726   lyGluLysGlyMetAlaGlyGlyHisGlyProAspGlyProLysGlySer  742

1214  .......................TGCCCGGTCAGCTCTGGGC  1195
         |||                   |||     |||:::
743   ProGlyProSerGlyThrProGlyAspThrGlyProProGlyLeuGlnGl  759

1194  ACGGCCTGGTACA..........GCCCCTCGCCCACGAAATCCGT  1160
         :||||||             |||||::|||
759   yMetProGlyGluArgGlyIleAlaGlyLyThrProGlyProLysGlyAsp.  775
```

```
1159  GTAAAACAGCGTGAGGTCATGAGTGCCATCCAGTCGCACAGCTCAGCCA  1110
         ...................................................  775

1109  CGAAGAGCCGGCCAGGGTGCGGGGCATGCGGCAGCACAGCTGGTGCAGC  1060
         |||::::|||    |||  ||||||     :::::|||:::||||||
776   ...ArgGlyGlyIleGlyIleGluLysGlyAlaGlyIleGlyThrAlaGlyAsnAs  791

1059  CGGGGAAGCAGGCGCCCAGGTTCCGGAAAGCCAAGCGGGCCCGGCATGG  1010
         |||          ::::||| ||||||       ||| |||||
791   pGlyAlaGlyLysGlyLeuProGly.........ProLeuGlyPro....  802

1009  ACAGCAGTGGGCGGCCACAAGGACGGGGGCCCAGCCCTTCGTGCGGTCGG  960
         |||||||:::|||:::|||||||
803   ...................ProGlyProAlaGlyLeuLeu.........  809

959   TGGGGCCCAGCCGCTGCCTCCTCAGCCACCAGCAGTGGCTGCTACGCAG  910
         |||    ||
810   ...GlyGluLysGlyGluProGlyProArg..................  818

909   GTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGAGCCACTCCT.......  870
         ||| |||    |||||||||||::::|||
819   ....GlyLeuValGlyProProGlySerArgGlyAsnProGlySerArgG  834

870   ...................................................  870

834   lyGluAsnGlyProThrGlyValAlaValGlyPheAlaGlyProGlnGlySer  850

869   ........CCTGGGTGCCCAGTAGGGGGCCAGGGCACTGTGTCCCAG  829
         |||||    |||:::  ||||||::  |||     :::  |||:::
851   AspGlyGlnProGlyValValProGlyGlu....ProGlyGlu.....ProGl  864

828   TCAATGCAGCAGGAGAGTAGCCCAGCAGCCCCCAAGACTGATCATGAA  779
         :|||     |||:::|||||||:::::||||:::::||||||
864   yGlnLysGlyAspAlaGlySerProGly..ProGlnGly..........  876

778   GGCCATAGAACAGAGTAGGCCTGGCGACAGTGTCCGGGT.........  741
         :::   ||||||        |||||| :::||
877   ....LeuAlaGlySerProGlyProHisGlyProAsnGlyValProGlyGly  891

740   .......................................CC  739
         ||
892   LeuLysGlyGlyArgGlyThrGlnGlyProGlyProGlyAlaThrGlyPhePr  908

738   CGGAAGAGGTCAGAGAGCAGGAGCCTCCAGTGGAGTGAAGCACACCTGGCC  689
         |||  :::  |||:::||||||     :::::||||:::||
908   oGlySerAlaGlyArgValGlyProProGlyProAlaGlyAlaProGly.  924

688   ACAGAAGTCCAGCAGCCCA.....CGCCCAGGATGAGCAGTGCCAGCT  645
         ::  ||||||        ::|||||||
925   ......ProAlaGlyProLeuGlyProGlyLysGlyGluGlyProPro  938

644   CCAGGGGCCTGGGATCCCGGGCACCAGCAGCCCTGCTAGCCAGCGGCGCCCTT  595
         ||||||||  ::: ||||||||     :::
939   GlyProArgGlyAspProGlySerHisGlyArgValGly...........  951

594   GGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAA  545
         ||| :::  :::||||||      |||||||:::
952   ........ValArgGlyProAlaGlyProProGly...............  960

544   GGGCCCGGCGGCGGGCCATAGCCGTCCCAGTGGTCACTGCTGCTGAGCCTA  495
         |||||:::|||                 ||||||
960   ...................................................  960

494   GGAGCGGGACACAGACCAGGCCACTGGACCAATGCCAGCACCATG  445
         |||   :::      :::|||||||:::::|||
961   GlyProGlyAspLysGlyAspProGlyGluAspGlyGlnProGlyProAs  977

444   GTCATGAACTTCTCCTCACCCCACTTCCAGCAGCAGAGGCGGCCACATA  395
         :::                  :::||||||::::||
977   pGly........................ProProGlyProAlaGlyThrGlyGlnA  989
```

```
 394 GGTGATGCCTGCGGCCAAACACCACCTCCAGGCCAAAGG............  357
     ||                         ||||||||||||
 989 rgGlyIleVal........GlyMetProGlyGlnArgGlyGluArgGly 1002

 356 ..........TTAGCAGTTGACCAGCAAGAGAGCTGGGCTTC  325
            |||||  |||||||||  |||||||:||| ::
1003 MetProGlyLeuProGlyProAlaGlyThrProGlyLysValGlyProTh 1019

 324 CGGTGCCGCAGCAGGC...............GGCTCACCCACAG  296
     :||||||||:::::                  |||  ::
1019 rGlyAlaThrGlyAspLysGlyProProValGlyProProGlyS 1036

 295 CCTCTGGACCATAGTGGG...CCAGGCGGGTAGGGCTCAGGGGGCCGTTC  249
     ::::||||||  |||  |||||  ||  :::||||||
1036 erAsnGlyPro.ValGlyLeuProGlyProGluGlyProAlaGlyAsnAs 1052

 248 AGGCACTCCAGAAACTGCTTCGTCTCGGCTCGTCCAGAAGCTGCGGCCT  199
     :||||||||    |||  |||||||  |||     |||
1052 pGlyThrProGly......ArgAspGlyAlaValGlyGluArgGly..... 1065

 198 CTCCTCCTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTTC  149
     :::::|||||                 :|||||
1066 ..............AspArgGlyAspProGly......... 1071

 148 TGCCAGCCCTTTGGTGCCGGTCCAGCTTCTCAGCCCATGCTGCCAACACCTG   99
     ||| |||        ::  :::  :||||||:
1072 ......ProAlaGlyLeuProGlySerGln.GlyAlaProGlyThrProG 1086

  98 CTGCTGTGGGGCACCTCAGTGGGGACACGTCTCATCACTCAGATCCTGGC   49
     ::  ||||||              |||||||:|||||||||
1086 lyProValGly...AlaProGlyAspAlaGlyGlnArgGlyAspProGly 1101

  48 ...........CAGGCCGCGGCGTGTCACC   29
             |||  ||||| ||
1102 SerArgGlyProIleGlyHisGlyLeuGlyArgAlaGlyLysArgGlyLeuPr 1118

  28 CGGAGCCAGCGCGTGCAGG   10
     ||||  ::::::  |||
1118 oGlyProGlnGlyProArg 1124
```

seq_name: SwissProt_37:CA12_HUMAN

seq_documentation_block:
```
ID   CA12_HUMAN     STANDARD;      PRT;   1418 AA.
AC   P02458;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR [CONTAINS: CHONDROCALCIN].
GN   COL2A1.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90067946.
RA   SU M.W., LEE B., RAMIREZ F., MACHADO M., HORTON W.;
RT   "Nucleotide sequence of the full length cDNA encoding for human type
RT   II procollagen.";
RL   NUCLEIC ACIDS RES. 17:9473-9473(1989).
RN   [2]
RP   SEQUENCE OF 1-28 FROM N.A.
RX   MEDLINE; 87031574.
RA   NUNEZ A.M., KOHNO K., MARTIN G.R., YAMADA Y.;
RT   "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL   GENE 44:11-16(1986).
RN   [3]
RP   SEQUENCE OF 432-1145 FROM N.A.
RA   RAMIREZ F.;
RL   SUBMITTED (DEC-1988) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE OF 963-1418 FROM N.A.
RX   MEDLINE; 85190534.
RA   CHEAH K.S.E., STOKER N.G., GRIFFIN J.R., GROSVELD F.G., SOLOMON E.;
RT   "Identification and characterization of the human type II collagen
RT   gene (COL2A1).";
RL   PROC. NATL. ACAD. SCI. U.S.A. 82:2555-2559(1985).
RN   [5]
RP   SEQUENCE OF 1120-1398 FROM N.A.
RX   MEDLINE; 85306861.
RA   ELIMA K., MAEKELAE J.K., VUORIO T., KAUPPINEN S., KNOWLES J.,
RA   VUORIO E.;
RT   "Construction and identification of a cDNA clone for human type II
RT   procollagen mRNA.";
RL   BIOCHEM. J. 229:183-188(1985).
RN   [6]
RP   SEQUENCE OF 1106-1418 FROM N.A.
RX   MEDLINE; 88067771.
RA   ELIMA K., VUORIO T., VUORIO E.;
RT   "Determination of the single polyadenylation site of the human pro
RT   alpha 1(II) collagen gene.";
RL   NUCLEIC ACIDS RES. 15:9499-9504(1987).
RN   [7]
RP   SEQUENCE OF 1227-1289 FROM N.A.
RX   MEDLINE; 86104139.
RA   NUNEZ A.M., FRANCOMANO C., YOUNG M.F., MARTIN G.R., YAMADA Y.;
RT   "Isolation and partial characterization of genomic clones coding for
RT   a human pro-alpha 1 (II) collagen chain and demonstration of
RT   restriction fragment length polymorphism at the 3' end of the gene.";
RL   BIOCHEMISTRY 24:6343-6348(1985).
RN   [8]
RP   SEQUENCE OF 1176-1226 FROM N.A.
RX   MEDLINE; 84118798.
RA   STROM C.M., UPHOLT W.B.;
RT   "Isolation and characterization of genomic clones corresponding to
RT   the human type II procollagen gene.";
RL   NUCLEIC ACIDS RES. 12:1025-1038(1984).
RN   [9]
RP   SEQUENCE OF 35-167 FROM N.A.
RX   MEDLINE; 89233138.
RA   SU M.W., BENSON-CHANDA V., VISSING H., RAMIREZ F.;
RT   "Organization of the exons coding for pro alpha 1(II) collagen N-
RT   propeptide confirms a distinct evolutionary history of this domain of
RT   the fibrillar collagen genes.";
RL   GENOMICS 4:438-441(1989).
RN   [10]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 91184577.
RA   KUIVANIEMI H., TROMP G., PROCKOP D.J.;
RT   "Mutations in collagen genes: causes of rare and some common diseases
RT   in humans.";
RL   FASEB J. 5:2052-2060(1991).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 97255959.
RA   KUIVANIEMI H., TROMP G., PROCKOP D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   HUM. MUTAT. 9:300-315(1997).
RN   [12]
RP   VARIANT SER-1074.
RX   MEDLINE; 90036909.
RA   VISSING H., D'ALESSIO M., LEE B., RAMIREZ F., GODFREY M.,
RA   HOLLISTER D.W.;
RT   "Glycine to serine substitution in the triple helical domain of pro-
RT   alpha 1 (II) collagen results in a lethal perinatal form of short-
RT   limbed dwarfism.";
RL   J. BIOL. CHEM. 264:18265-18267(1989).
RN   [13]
RP   VARIANT SEDC GLY-1095--TYR-1330 DEL.
RX   MEDLINE; 89266907.
RA   LEE B., VISSING H., RAMIREZ F., ROGERS D., RIMOIN D.;
RT   "Identification of the molecular defect in a family with
RT   spondyloepiphyseal dysplasia.";
RL   SCIENCE 244:978-980(1989).
```

RN   [14]
RP   VARIANT OSTEOARTHRITIS CYS-650.
RX   MEDLINE; 90370826.
RA   ALA-KOKKO L., BALDWIN C.T., MOSKOWITZ R.W., PROCKOP D.J.;
RT   "Single base mutation in the type II procollagen gene (COL2A1) as a
RT   cause of primary osteoarthritis associated with a mild
RT   chondrodysplasia.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:6565-6568(1990).
RN   [15]
RP   VARIANT OI-IV VAL-717.
RX   MEDLINE; 91291136.
RA   BATEMAN J.F., HANNAGAN M., CHAN D., COLE W.G.;
RT   "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT   valine substitution in osteogenesis imperfecta type IV. Detection of
RT   the mutation and prenatal diagnosis by a chemical cleavage method.";
RL   BIOCHEM. J. 276:765-770(1991).
RN   [16]
RP   VARIANT OSTEOARTHRITIS CYS-650.
RX   MEDLINE; 91086471.
RA   EYRE D.R., WEIS M.A., MOSKOWITZ R.W.;
RT   "Cartilage expression of a type II collagen mutation in an inherited
RT   form of osteoarthritis associated with a mild chondrodysplasia.";
RL   J. CLIN. INVEST. 87:357-361(1991).
RN   [17]
RP   VARIANT HYPOCHONDROGENESIS GLU-984.
RX   MEDLINE; 93054548.
RA   BOGAERT R., TILLER G.E., WIES M.A., GRUBER H.E., RIMOIN D.L.,
RA   COHN D.H., EYRE D.R.;
RT   "An amino acid substitution (Gly853->Glu) in the collagen alpha
RT   1(II) chain produces hypochondrogenesis.";
RL   J. BIOL. CHEM. 267:22522-22526(1992).
RN   [18]
RP   VARIANT HYPOCHONDROGENESIS SER-705.
RX   MEDLINE; 92262484.
RA   HORTON W.A., MACHADO M.A., ELLARD J., CAMPBELL D., BARTLEY J.,
RA   RAMIREZ F., VITALE E., LEE B.;
RT   "Characterization of a type II collagen gene (COL2A1) mutation
RT   identified in cultured chondrocytes from human hypochondrogenesis.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 89:4583-4587(1992).
RN   [19]
RP   VARIANT WS-II ASP-198.
RX   MEDLINE; 93304428.
RA   KORKKO J., RITVANIEMI P., HAATAJA L., KAARIAINEN H., KIVIRIKKO K.I.,
RA   PROCKOP D.J., ALA-KOKKO L.;
RT   "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT   for glycine alpha 1-67 and that causes cataracts and retinal
RT   detachment: evidence for molecular heterogeneity in the Wagner
RT   syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL   AM. J. HUM. GENET. 53:55-61(1993).
RN   [20]
RP   VARIANT SEMD CYS-940.
RX   MEDLINE; 93252400.
RA   TILLER G.E., WEIS M.A., LACHMAN R.S., COHN D.H., RIMOIN D.L.,
RA   EYRE D.R.;
RT   "A dominant mutation in the type II collagen gene (COL2A1) produces
RT   spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";
RL   AM. J. HUM. GENET. 53:A209-A209(1993).
RN   [21]
RP   VARIANT OSTEOARTHRITIS CYS-650.
RX   MEDLINE; 93282819.
RA   HOLDERBAUM D., MALEMUD C.J., MOSKOWITZ R.W., HAQQI T.M.;
RT   "Human cartilage from late stage familial osteoarthritis transcribes
RT   type II collagen mRNA encoding a cysteine in position 519.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 192:1169-1174(1993).
RN   [22]
RP   VARIANT SEMD ARG-285.
RX   MEDLINE; 93252400.
RA   VIKKULA M., RITVANIEMI P., VUORIO A.F., KAITILA I., ALA-KOKKO L.,
RA   PELTONEN L.;
RT   "A mutation in the amino-terminal end of the triple helix of type II
RT   collagen causing severe osteochondrodysplasia.";
RL   GENOMICS 16:282-285(1993).
RN   [23]
RP   VARIANT SEDC CYS-206.

RX   MEDLINE; 94063862.
RA   WILLIAMS C.J., CONSIDINE E.L., KNOWLTON R.G., REGINATO A., NEUMANN G.,
RA   HARRISON D., BUXTON P., JIMENEZ S.A., PROCKOP D.J.;
RT   "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT   family with an Arg75-->Cys mutation in the procollagen type II gene
RT   (COL2A1).";
RL   HUM. GENET. 92:499-505(1993).
RN   [24]
RP   VARIANT SEDC CYS-920.
RX   MEDLINE; 9331508.
RA   CHAN D., TAYLOR T.K.F., COLE W.G.;
RT   "Characterization of an arginine 789 to cysteine substitution in
RT   alpha 1(II) collagen chains of a patient with spondyloepiphyseal
RT   dysplasia.";
RL   J. BIOL. CHEM. 268:15238-15245(1993).
RN   [25]
RP   VARIANT SEDC SER-1128.
RX   MEDLINE; 93140139.
RA   COLE W.G., HALL R.K., ROGERS J.G.;
RT   "The clinical features of spondyloepiphyseal dysplasia congenita
RT   resulting from the substitution of glycine 997 by serine in the alpha
RT   1(II) chain of type II collagen.";
RL   J. MED. GENET. 30:27-35(1993).
RN   [26]

```
alignment_scores:
          Quality:    325.00                  Length:    848
          Ratio:        0.895                  Gaps:       51
Percent Similarity:    42.807   Percent Identity:    26.887

alignment_block:
US-09-030-606-110/rev x CA12_HUMAN

Align seg 1/1 to: CA12_HUMAN  from: 1  to: 1418

2164 CAGTCAGGCAGCCCTTAGAGACTGGGAGGAGGAGGAGGACGCCCAGCC 2115
     |||::||||:::||| ||| |||  :::  |||||||||| ||
  36 GlnLysGlyGluProGlyArgIleLysAspIleLysValGlyProLysGlyPr 52

2114 CCCAGCTGTGCAGCTACGCACCTCAGCAGCACAGGGTGCAGCAGAGAGC 2065
     ||||:||  ||||              :::|||:::::::  |
  52 oProGlyProGln...........GlyProAlaGlyGluGlnGlyP 64

2064 CACATTACTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCCAT 2015
     ||:|:  ::|||:||:::::  :::  ||| ||::  |||
  64 roArgGlyAspArgGlyAspLysGlyGluLysGlyAlaProGlyProArg 80

2014 GGGGCTAACAGGAGCGGGGAGCTGGGA................CC 1986
     ||  |||::||||||  |||                      ||
  81 Gly......ArgAspGlyGluProGlyThrLeuGlyAsnProGlyProPr 95

1985 CAGTGAGGCAGGCCCT......CCACCCCAATGTGCTGGAAGTTTTCTAC 1942
     ||::      ||||||       ::::::|||
  95 oGlyProProGlyProProGlyProProGlyLeuGlyGlyAsnPheAlaA 112

1941 GCTGAGTATTTGGCCAAGTCGCCTCTTGTCAAATACTACCTGTGTAGCAAA 1892
     ::: ||                         :: ::::||
 112 laGlnMetAlaGlyGly.....................PheAspGluLys 121

1891 GTAAATGGCGACCAGAGACCTCGGGCAGCAGACACCATATAGGCAGTGA 1842
     :::::|||  |||  ||| |||:: ||| ||| |||
 122 AlaGlyGlyAlaGlnLeu..GlyValMetGlnGlyPro........... 133

1841 CAGACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATGGGGCCACC 1792
     |||  :||::      ::|||  |||||||
 134 ..........MetGlyProMetGlyPro.........ArgGlyProPr 143

1791 TGGGACAGCAGGAAGGCAGCTATCCA...GGATGGCGAGGT......... 1755
     |||  ||||||  |||||||:::::::||
 143 oGlyProAlaGlyAlaProGlyProGlnGlyPheGlnGlyAsnProGlyG 160

1754 ..CCAGGCAGATGCCCCGGCCCGGAACCACCCTGG...CCTCGGTGGGGCT 1710
     |||||||::::  |||  :::||| ||| |||||
```

```
160  luProGlyGlu...ProGlyValSerGlyProMetGlyProArgGlyPro 175

1709 CACCCCACCACA...CGTACGGAGACATCACAGGCAGGAGGCCCCGCA 1664
     ||||||                  :::|||
176  ProGlyProProGlyLysProGlyAspAspGlyGluAlaGlyLysProGl 192

1663 GAGCGCGGTGAGGGTGGGAGCAGGCCACTGCCTCCAGCACCCACGTGTC 1614
     :::||||      |||          ||||||:::|||
192  yLysAlaGlyLeuArgGly.........ProProGlyProGlnGlyA 205

1613 CATTAGGG...AAGGGAGCTCCAGGCTTAGGGCCTGGCAGGAAGCTGGTC 1567
     |||      |||:::|||||||| ||| ||||
205  laArgGlyPheProGlyThrProGlyLeu...ProGlyValLys..GlyH 220

1566 ATCAGGCTGTCCACTGCTAGCACCTCCAGTGTCCCCTCGGTATTTGGG 1517
     ||:::|||      |||||:::
220  isArgGly.................................... 222

1516 CAGGAACACCTGCTTCTCCCGGTGGTAGAGGA......GGCCAGTGTGT 1473
     |||              :::::|||
223  ...............TyrProGlyLeuAspGlyAlaLysGlyGluAlaGlaGl 234

1472 AGGGCAGGATCTGCAGGGCTGAGAGGTGAACCCGGTGAGGGCGGCTGAA 1423
     :::      :::|||:::||||||||:::|||
234  yAlaProGlyValLysGlyGluSerGlySerProGlyGluAsnGly.... 249

1422 GCTGTCACCACGGCCACACTGTGGGACAGGCATGTGCACCGCAGCCAC 1373
     :::|||||||
250  .......SerProGlyPrometGlyProArgGlyProArgGlyLeuPro 260

1372 AGGGAAAGCTGCCACACTGTCTCCGGGCCAAATAGACTGCTCGAGTGCCGAATCGCT 1324
     |||:::      ::::|||:::::||||||:::
261  GlyGluArgGlyArgThrGlyAlaProGlyAlaLysGlyAlaAlaArgGlyAs 277

1323 ..........GCACCAGCGCGGTCCATGACCAGAGAGAAGACCAGGGA 1286
             :::||||||||||
277  nAspGlyGlnProGlyProAlaGlyPro.............ProGlyP 289

1285 GATGGGCGCACTGCAGGAACAGCC......CCAGGCTGCCCATCCGAACGC 1242
     |||||:::        ||||||
289  roValGlyProAlaGlyGlyGlyProGlyPheProGlyAlaProGlyAlaLys 305

1241 CTTCATCATAGTGTCTCCGGGCCGCCTCCGGTCCGCCGGCCTCAGCCTCTGGGCACG 1192
     |||      :::|||   |||  |||
306  GlyGluAlaGlyProThrGlyArgThrGlyAlaArgGlyProGluGlyAlaGlnGlyPr 322

1191 CCCTGGTACAGCCCCCTCGCCCACGAAATCCGTGTAAAACAGCGTGAAGGT 1142
     |||:::||||||        |||
322  oArgGlyGluProGlyThrProGlySerPro................ 332

1141 CATGATGCCATCCAGCTGCACAGCTCAGCCACGAAGAGAGCCGGCGCCAGGG 1092
     |||||:::
333  ....GlyProAlaGlyAlaSerGlyAsnProGlyThrAspGlyIlePro 347

1091 TGCGGGGCATGCGGCCACAGCAGCTGTGCAGCCGGGGAAGCAGGAGGCGCCC 1042
     :::|||||  ||||||||  ||| :::|||
348  ...GlyGluAlaLysGlySer..AlaGlyAlaProGlyIleAlaGlyAlaPr 362

1041 AGGTTCCGGAAAGCCAAGCGGGCCCCGCCATGACAGCAGTGGGGGACA 993
     |||     :::||||||||||
362  oGlyPheProGlyAlaProGlyArgGly...ProAspProGlnGlyAlaThrG 378

993  ..................GGCTCACCCAAGCCCTGGA.............. 993
                    |||||:::|||
378  lyProLeuGlyProLysGlyGlnThrGlyLysProGlyIleAlaGlyPhe 394

992  ..............AGAGGGGGCCGACAGCCCTTCTGCTGCTGGGCCCCAGC 949
               :::|||||  |||||| ||| ||||||
395  LysGlyLysGlyGlnGlyProLys..........GlyGluProGlyProAl 407

948  GCTGCCTCCT.......... 939
     |||  |||
407  aGlyProGlnGlyAlaProGlyProAlaGlyProAlaGlyGluGlyGluGlyLysArgGlyA 424
```

```
938  ..........CAGCCACCAGCAGTGTGGCTGCTACGCAGTGAGGAAGATGA 897
                ::|||    ::|||  |||:::
424  laArgGlyGluProGlyGlyValGlyProIleGlyProGlyGluArg 440

896  GGGTGAGCAGGCCAAAGAGGC........................ACTCC 871
     |||     |||:::|||
441  GlyAlaProGlyAsnArgGlyPheProGlyGlnAspGlyLeuAlaGlyPr 457

870  TCCTGGGTGCCCAGGT...AGGGGGCCAGGGCACTGGTGTCCCAGTCAAT 824
     |||:::||||||      |||||||
457  oLysGlyAlaProGlyLeuArgGlyGluArgGlyProSerGlyLeuAlaGlyProLysG 474

823  GGCAGGCAGGA......GGTAGCCCAGGCAGCCCCCAAGACTGATCATGA 780
     |||          |||:::|||
474  lyAlaAsnGlyAspProGlyArgProGlyArgProGlyGluProGlyAla 490

779  AGGCATAGACAGAGTAGGCCTGGCGACAGTGGTCCGGGTCCCGGACGAGG 730
     |||:::||||||||::::|||
491  ArgGlyLeuThrGlyArgProGlyTyrAspPAlaGlyPro..GlnGlyLysVa 506

729  TCAGAGAGCAGGGCCTCCAGTGGAGTGAAGCACACTGGCCACACAGAAGTC 680
     ::||| |||  ||| :::|||  |||||
506  lGlyProSerGlyAlaProGlyLeuAspGlyArgProGlyProProGlyP 523

679  CAGCAGCCCACCC.............CCAGGATGAGCA 654
     || :::  |||              |||||
523  roGlnGlyAlaArgGlyGlnProGlyValMetGlyPheProGly...... 537

653  GTGTCCAGCTCCAGGGGCCTGGGATCCGGGCCACAGCAGCCCTGCTAGCCAG 604
     |||    :::|||||          |||:::
538  ...ProLysGlyAlaAsnGlyGluProGlyLysAlaGly....... 549

603  CCGGCCCTGGGATGAGAAAGAGGCTCAGCAGGATGCCA......AGGA 560
     |||||| :::|||    ||||||
550  ......GluLysGlyLeuProGlyAlaProGlyLeuArgG 561

559  CAGTGCCCAGATGAAGGGCCGGCGGCCGCCATAGCGTCCACGCCAGTGGT 510
     :::|||||                        |||
561  .....................................Gly 564

509  CACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCAGCA.......... 468
     |||                            |||||||||
565  LysAspGlyGluThrGlyAlaLysGlyAspGlyProGlyProAlaGlyProAl 581

...  ..............CTGGACCAATGCCCAGCACCATGGTCATGAACTTCTC 431
                   |||:::||| |||
581  aGlyGluArgGlyGluGlnGlyAlaProGlyPro.........SerG 594

430  CTCTACCCCCACTTCCAGCACGAGAGGCGGCACATAGTGATGCCTCGG 381
     ||||||||
594  lyPheGlnGlyLeuProGly.................... 600

380  CCAAACACACCTCCAGCCCAAAGGTTAGCAGGTTGACCAGCAAGAGCTGG 331
     ::::|||||||||||
601  ...GlyLeuLysProGlyAspGlnGl 615

330  GCTTTCCGGTGCCCAGCAGGC................. 309
     |||:::  |||:::|||
615  yValProGlyGluAlaGlyAlaProGlyLeuValGlyProArgGlyGluA 632

308  ..................GGCTCACCCAAGCCCTGGA............ 288
                    ||||||:::|||
632  rgGlyPheProGlyGluGlySerProGlyAlaGlnGlyLeuGlnGly 648

287  .........CCATAGTGGCCGCCAAAGGTGGCGGGGTAGGGGGCCGTT 250
              ::::||| |||
649  ProArgGlyLeuPro.GlyThrProAspGlyLysProGlyAlaS 665

249  CAGGCACTCCAGAACTGCTTCGTCTGCTCCGGCTGTCTGCTCCAGAAGCTGGGCC 200
     |||||
665  erGlyPro..............AlaGlyProProGlyAlaGln 674
```

```
199 TCTCCTCCTTGCTGCGCGCAACTGCCTAGGAATCAGCCAGGCGCCCATT 150
        ::||||||    ::: ::||||  ::: :||||  ::||||  ::|||
675 GlyProProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyIl 691

149 CTGCCAGCCCTTGGT................................GCCGGTCCCAGCTT 121
        ::: ||| |||                                  |||||
691 eAlaGlyProLysGlyAspArgGlyAspArgProValGlyGluLysGlyProGluG 708

120 CTCAGCCC...........................ATGCTCAACACC 101
        |||                              :||::: ||
708 lyAlaProGlyLysAspGlyGlyArgGlyLeuThrGlyProIleGly.Pr 724

100 TGCTGCTGTGGGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTG 51
    |     ::: :||||::: :||||:: ::::  :::   |||
724 oProGlyProAlaGlyAlaAsnGlyGluLysGlyGluValGlyProProG 741

 50 GC...........CGAGGCGCGGCGCTGTCACCCGGA 25
    ||              |||||||||| ::|:||||
741 lyProAlaGlySerAlaGlyAlaAlaArgGlyAlaProGly 753
```

seq_name: SwissProt_37:CA54_HUMAN

seq_documentation_block:
```
ID   CA54_HUMAN     STANDARD;      PRT;   1685 AA.
AC   P29400; Q16126;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR.
GN   COL4A5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94165049.
RA   ZHOU J., LEINONEN A., TRYGGVASON K.;
RT   "Structure of the human type IV collagen COL4A5 gene.";
RL   J. BIOL. CHEM. 269:6608-6614(1994).
RN   [2]
RP   SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 92316923.
RA   ZHOU J., HERTZ J.M., LEINONEN A., TRYGGVASON K.;
RT   "Complete amino acid sequence of the human alpha 5 (IV) collagen
RT   chain and identification of a single-base mutation in exon 23
RT   converting glycine 521 in the collagenous domain to cysteine in an
RT   Alport syndrome patient.";
RL   J. BIOL. CHEM. 267:12475-12481(1992).
RN   [3]
RP   SEQUENCE OF 85-1685 FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 90337990.
RA   PIHLAJANIEMI T., POHJOLAINEN E.R., MYERS J.C.;
RT   "Complete primary structure of the triple-helical region and the
RT   carboxyl-terminal domain of a new type IV collagen chain, alpha
RT   5(IV).";
RL   J. BIOL. CHEM. 265:13758-13766(1990).
RN   [4]
RP   SEQUENCE OF 924-1685 FROM N.A.
RX   MEDLINE; 91169491.
RA   ZHOU J., HOSTIKKA S.L., CHOW L.T., TRYGGVASON K.;
RT   "Characterization of the 3' half of the human type IV collagen alpha
RT   5 gene that is affected in the Alport syndrome.";
RL   GENOMICS 9:1-9(1991).
RN   [5]
RP   SEQUENCE OF 914-1685 FROM N.A.
RX   MEDLINE; 90160375.
RA   HOSTIKKA S.L., EDDY R.L., BYERS M.G., HOEXHTYAE M., SHOWS T.B.,
RA   TRYGGVASON K.;
RT   "Identification of a distinct type IV collagen alpha chain with
RT   restricted kidney distribution and assignment of its gene to the
RT   locus of X chromosome-linked Alport syndrome.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:1606-1610(1990).
RN   [6]
RP   SEQUENCE OF 1442-1471 FROM N.A.
RX   MEDLINE; 90252791.
RA   MYERS J.C., JONES T.A., POHJOLAINEN E.R., KADRI A.S., GODDARD A.D.,
RA   SHEER D., SOLOMON E., PIHLAJANIEMI T.;
RT   "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
RT   to the region of the X chromosome containing the Alport syndrome
RT   locus.";
RL   AM. J. HUM. GENET. 46:1024-1033(1990).
RN   [7]
RP   SEQUENCE OF 1-20 FROM N.A.
RA   GUO C., VAN DAMME B., VANRENTERGHEM Y., DEVRIENDT K., CASSIMAN J.,
RA   MARYNEN P.;
RL   SUBMITTED (SEP-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [8]
RP   SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM)
RX   MEDLINE; 94133540.
RA   GUO C., VAN DAMME B., VAN DAMME-LOMBAERTS R., VAN DEN BERGHE H.,
RA   CASSIMAN J.-J., MARYNEN P.;
RT   "Differential splicing of COL4A5 mRNA in kidney and white blood
RT   cells: a complex mutation in the COL4A5 gene of an Alport patient
RT   deletes the NC1 domain.";
RL   KIDNEY INT. 44:1316-1321(1993).
RN   [9]
RP   REVIEW ON VARIANTS.
RX   MEDLINE; 97338662.
RA   LEMMINK H.H., SCHROEDER C.H., MONNENS L.A.H., SMEETS H.J.M.;
RT   "The clinical spectrum of type IV collagen mutations.";
RL   HUM. MUTAT. 9:477-499(1997).
RN   [10]
RP   VARIANT AS SER-1564.
RX   MEDLINE; 91169492.
RA   ZHOU J., BARKER D.F., HOSTIKKA S.L., GREGORY M.C., ATKIN C.L.,
RA   TRYGGVASON K.;
RT   "Single base mutation in alpha 5(IV) collagen chain gene converting a
RT   conserved cysteine to serine in Alport syndrome.";
RL   GENOMICS 9:10-18(1991).
RN   [11]
RP   VARIANT AS ARG-325.
RX   MEDLINE; 92303559.
RA   KNEBELMANN B., DESCHENES G., GROS F., HORS M.-C., GRUNFELD J.-P.,
RA   TRYGGVASON K., GUBLER M.-C., ANTIGNAC C.;
RT   "Substitution of arginine for glycine 325 in the collagen alpha 5
RT   (IV) chain associated with X-linked Alport syndrome: characterization
RT   of the mutation by direct sequencing of PCR-amplified lymphoblast
RT   cDNA fragments.";
RL   AM. J. HUM. GENET. 51:135-142(1992).
RN   [12]
RP   VARIANT AS GLU-325.
RX   MEDLINE; 93244772.
RA   RENIERI A., SERI M., MYERS J.C., PIHLAJANIEMI T., MASSELLA L.,
RA   RIZZONI G., DE MARCHI M.;
RT   "De novo mutation in the COL4A5 gene converting glycine 325 to
RT   glutamic acid in Alport syndrome.";
RL   HUM. MOL. GENET. 1:127-129(1992).
RN   [13]
RP   VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX   MEDLINE; 94010948.
RA   LEMMINK H.L., SCHROEDER C.H., BRUNNER H.G., NELEN M.R., ZHOU J.,
RA   TRYGGVASON K., HAGGSMA-SCHOUTEN W.A.G., ROODVOETS A.P., RASCHER W.,
RA   VAN OOST B.A., SMEETS H.J.M.;
RT   "Identification of four novel mutations in the COL4A5 gene of
RT   patients with Alport syndrome.";
RL   GENOMICS 17:485-489(1993).
RN   [14]
RP   VARIANTS AS E-400;V-406;V-638;A-638;R-653;R-796;R-869;R-872 & C-1241.
RX   MEDLINE; 95329276.
RA   BOYE E., FLINTER F., ZHOU J., TRYGGVASON K., BOBROW M., HARRIS A.;
RT   "Detection of 12 novel mutations in the collagenous domain of the
RT   COL4A5 gene in Alport syndrome patients.";
RL   HUM. MUTAT. 5:197-204(1995).
RN   [15]
```

RP   VARIANT AS ARG-1649.
RX   MEDLINE; 96213750.
RA   BARKER D.F., PRUCHNO C.J., JIANG X., ATKIN C.L., STONE E.M.,
RA   DENISON J.C., FAIN P.R., GREGORY M.C.;
RT   "A mutation causing Alport syndrome with tardive hearing loss is
RT   common in the western United States.";
RL   AM. J. HUM. GENET. 58:1157-1165(1996).
RN   [16]
RP   VARIANTS AS.
RX   MEDLINE; 96213754.
RA   RENIERI A., BRUTTINI M., GALLI L., ZANELLI P., NERI T., ROSSETTI S.,
RA   TURCO A., HEISKARI N., ZHOU J., GUSMANO R., MASSELLA L., BANFI G.,
RA   SCOLARI F., SESSA A., RIZZONI G., TRYGGVASON K., PIGNATTI P.F.,
RA   SAVI M., BALLABIO A., DE MARCHI M.;
RT   "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT   exons of the COL4A5 gene.";
RL   AM. J. HUM. GENET. 58:1192-1204(1996).
RN   [17]
RP   VARIANT AS ASP-1498.
RX   MEDLINE; 96233932.
RA   TVERSKAYA S., BOBRYNINA V., TSALYKOVA F., IGNATOVA M.,
RA   KRASNOPOLSKAYA X., EVGRAFOV O.;
RT   "Substitution of A1498D in noncollagen domain of a5(IV) collagen
RT   chain associated with adult-onset X-linked Alport syndrome.";
RL   HUM. MUTAT. 7:149-150(1996).
RN   [18]
RP   VARIANT AS GLN-1677.
RX   MEDLINE; 97295089.
RA   BARKER D.F., DENISON J.C., ATKIN C.L., GREGORY M.C.;
RT   "Common ancestry of three Ashkenazi-American families with Alport
RT   syndrome and COL4A5 R1677Q.";
RL   HUM. GENET. 99:681-684(1997).
RN   [19]
RP   VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
RX   MEDLINE; 98112435.
RA   NERI T.M., ZANELLI P., DE PALMA G., SAVI M., ROSSETTI S., TURCO A.E.,
RA   PIGNATTI G.F., GALLI L., BRUTTINI M., RENIERI A., MINGARELLI R.,
RA   TRIVELLI A., PINCIAROLI A.R., RAGAIOLO M., RIZZONI G.F., DE MARCHI M.;
RT   "Missense mutations in the COL4A5 gene in patients with X-linked
RT   Alport syndrome.";
RL   HUM. MUTAT. SUPPL. 1:S106-S109(1998).
-!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC   GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC   MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/
CC   NIDOGEN.
-!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC   ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC   WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
-!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC   DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC   G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC   CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC   TRIPLE-HELICAL 7S DOMAIN.
-!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC   UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
-!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC   ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC   THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC   IV COLLAGENS.
-!- DISEASE: DEFECTS IN COL4A5 ARE ASSOCIATED WITH X-LINKED ALPORT
CC   SYNDROME (AS). ALPORT SYNDROME IS CHARACTERIZED BY PROGRESSIVE
CC   GLOMERULONEPHRITIS, OFTEN ASSOCIATED WITH HIGH-TONE SENSORINEURAL
CC   DEAFNESS, SPECIFIC EYE ABNORMALITIES (LENTICONUS AND MACULAR
CC   FLECKS), AND GLOMERULAR BASEMENT MEMBRANE DEFECTS. IN MALES, THE
CC   TYPICAL TIME COURSE FOR THE PROGRESS OF ALPORT SYNDROME IS:
CC   HEMATURIA BY THE AGE OF 5 YEARS, DEAFNESS AND HYPERTENSION IN
CC   EARLY TEENAGE LIFE, DETERIORATION OF RENAL FUNCTION BY AGE 20, AND
CC   END-STAGE RENAL FAILURE SOON THEREAFTER. FEMALES TEND TO FOLLOW A
CC   MUCH MILDER COURSE AND RARELY GO INTO RENAL FAILURE.
-!- ALTERNATIVE PRODUCTS: TWO FORMS ARE PROBABLY PRODUCED BY
CC   ALTERNATIVE SPLICING OF THE SAME GENE. THE LONGER FORM, FOUND IN
CC   KIDNEY, INTRODUCES 2 EXTRA G-X-X REPEATS INTO THE TRIPLE-HELIX
CC   DOMAIN.

```
alignment_scores:
       Quality:   323.00        Length:   1109
       Ratio:     0.756         Gaps:     76
Percent Similarity: 38.503   Percent Identity: 25.609

alignment_block:
US-09-030-606-110 x CA54_HUMAN

Align seg 1/1  to: CA54_HUMAN   from: 1  to: 1685

   2 GGAACCAGCCTGCACGCCGCTGGCTCCGGGTGACAGCCGCGGCCCT.......  46
     ||||||::::: ::  :: ||||||||  |||
 486 GlyThrGlyIleSerGly...ProProGlnProGlyLeuProGlyLe     501

  47 ......CGGCCAGGATCTGAGTGATGAGACGTGTCCCCACTGAGGTGCC   89
       ||||||||||                                 |||
 501 uProGlyProProGlySer.........................LeuGlyPheP  511

  90 CCACAGCAGCAGGTGTTGAGCATGGCTGAGAAGCTGGACCGGCACCAAA  139
     |||  |||||  |||   :::|||||| |||  |||
 511 roGlyGlnLysGly...GluLysGlyInAlaGlyAlaThrGlyProLys  526

 140 GGGCTG..............GCAGAAATGGCGGCCTGGCTGATTCCTAG  174
     ||||||               |||     |||||||||||  |||
 527 GlyLeuProGlyIleProGlyAla.ProGlyAlaProGly...PheProG  542

 175 GCAGTTGCGGGCAGCAAGGA..............................  194
     ||||||                 ::  |||
 542 lySerLysGlyGluProGlyAspIleLeuThrPheProGlyMetLysGly  558

 195 ......GGAGAGGCCCAGCTTCTGGAGCAGCAGGCCGAGACGAAGCAGTTC  238
        |||||||  :::::||||||                       |||
 559 AspLysGlyGluLeuGlySerProGlyAlaPro.........GlyLeuPr  572

 239 TGGAGTGCCTGAACGGCCCCCTGAGCCTACCACCGTCCACCGGCCACTATGGT  288
     |||::||| |||  ::  |||   ||||||  ||||| |||
 572 oGlyLeuProGlyThrProGlyGlnAspGlyLeuProGly.LeuProGly  588

 289 CCAGAGGCGTGTGGGTGAGCCG............................  317
     ||::||  |||||||                           ::
 589 ProLys......GlyGluProGlyGlyIleThrLysGlyGluArgGl  603

 318 GGCACCGGAAAGCCCAGCTCTTGCTGTCAACCTGCTAACCTTTGGCCTG  367
     |||  :::||:::||| ||| |||||::|||
 603 yProProGlyAsnProGlyLeuProGlyLeuProGlyAsnIleGlyProM  620

 368 GAGGTGTGTTTGGCCCAGGCATCACCTATGTCCGCCTCTGCTGGA  417
     |||||| |                                 |||
 620 etGly....................................ProProGly  624

 418 AGTGGGGTAGAGGAGAAGTTCATGACCATGGT.........GCTGG  455
     |||          ||| ||| :: |||||
 625 PheGlyProProGlyProValGlyGlyIleGlnGlyIleGlyValAlaGl  641

 456 GCATTGGTCCAGTGCTGGGCCTGGTCTGTGTCCCGCTCCTAGGCTCAGCC  505
     |||::::|||||||:::: |||:::||:::
 641 y......AsnProGlyGlnProGlyIle...ProGlyProLys.......  652

 506 AGTGACCACTGGCGTGGACGTATGGCCGCCGGCCCTTCATCTGGGC  555
     |::||||||||||||||||||||||||||
 652 ................................................  652

 556 ACTGTCCTTGGGCATCCTGCTGAGCCTCTTTCTCATCCCAAGGGCCGCGCT  605
     |||  ||||:::::  ::::::::|||
 653 ........GlyAspProGlyGlnThrIleThrGlnPro.............  662

 606 GGCTAGCAGGGCTGCTGTGCCGGATCCCAGGCCCCTGGAGCTGGCACTG  655
     :::::: |||||||  :::::|||||||| :::  :::
```

```
663  ....GlyLysProGlyLeuProGlyAsnProGlyLyArgAspGlyAspValG 678

656  CTCATCCTGGG...........CGTGGGGCTGCTGGACTTCTGTGCCCA 693
     ::  ||||||                   ||||||            |||
678  lyLeuProGlyAspProGlyLeuGlnProGlyLeu.....Pro 692

694  GGTGTGCTTCACTCCACTGAGGACGCCCTGCTCTCTGACCT..........CT 734
     ||||:   |||  ||||:: ||  |||||:||         |||
693  GlyIleProGlySerLysGlyGluProGlyIle...ProGlyIleGlyLe 708

735  TCCGGGA...CCCGGACCACTCTCG...CCAGGCCTACTCTGTCTATGCC 778
     ||||||  ||||||||  :::  ||||||
708  uProGlyProProGlyProLysGlyProGlyPheProGlyIlePro....... 721

779  TTCATGATCAGTCTTGGGGCTGCCTGGGCTACCTTCCTGCCTGCCATTGA 828
     ||||||||                     :::
722  ........GlyProProGlyAlaProGlyThrProGlyLeuArg 732

829  CTGGGACACCAGTGCCCTGCCCCTGCCCTACCTGGCACCCAGGA....... 870
     :::||   |||     ||||||                  ||||||
733  IleGlyLeuGluGlyProProGlyProPheProGlyProLysGl 749

871  .......GGAGTGCCTCTTTGGCCTGCTGCCACCCTCATCTTCCTC 907
         :::::|||     |||:::  |||  |||||
749  yGluProGlyPheAlaLeuProGlyProProGlyProGlyLeuProG 766

908  ACCTGGCTAGCAGCCACACTGCGTGGTGGCGAGGAGGCAGCCGTGGGCCC 957
     ||::::||||||             |||  |||||
766  lyPheLysGlyAlaLeuGlyLysProGlyAspArgGlyPheProGlyPro 782

958  CACCGAGCCAGCAGAGGGCTGTCGCCCCCTCCTTGTCGCCCCACTGCT 1007
     |||||||                       :::  |||
783  .......ProGlyProProGlyArgThrGlyLeuAs 792

1008 GTCCATGCGGCCCCGCCTTGGCTTTCCGGAACCTGGGCGCCCTGCTTCCC 1057
     :::::|||||| ||::|                  |||::|
792  pGlyLeuProGlyProLysGlyAspValGlyProAsnGlyGlnProGlyP 809

1058 CGGCTGCCACCAGTCTGCTGCCG.........CATGCCCCGCACCCT 1095
     ||  ::|||                  ||||||
809  roMetGlyProProGlyLeuProGlyIleGlyValGlnGlyProProGly 825

1096 GCGCCGGCTCTTCGCTGGCTGAGCTGTGCAGCGTGATGATGCACTGACCT 1145
     |||||||    |||  |||            :::
826  ProProGlyIleProGlyProIleGlyGln.......... 835

1146 TCACGCTGTTTACACGGATTTCGTGGGCAGAGGGCTGTACCAGGGCGTG 1195
     ||||||||       ||||||:::::
836  ...ProGlyLeuHisGly..........IleProGlyGlu. 844

1196 CCCAGAGCTGAGCCGGGCACCGGGCCCGGAGACACTATGATGAAGGCGT 1245
     ||       ||||||                   |||
845  ........LysGlyAspProGlyProProGlyLeu. 853

1246 TCGGATGGGCAGCCTGGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGTCT 1295
     ||||||   |||  |||     ||||||::
854  ......AspValProGlyPro..ProGlyGluArgGlySerProGlyIl 867

1296 TCTCTCTGGTCATGGACCGGCTGGTGCAGGATTCGGCACTCGAGCAGTC 1345
     :  ::||||| |||||    |||  |||  |||
867  eProGlyAlaProGlyProIleGlyProPro.....GlySerProGlyL 882

1346 TATTTGGCCAGTGTGGCAGC....TTTCCCGTGCGTGCGGGTGCCAC 1389
     ||  ||::::||||::      ||||||  :::  :::
882  euProGlyLysAlaGlyAlaSerGlyPhePro...GlyThrLysGlyGlu 897

1390 ATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCCTCACCG 1439
     ||:   ||||| |||||||  |||
898  MetGlyMet.......................................... 900

1440 GGTTCACCTTCTCAGCCCTGCCGCAGATCCTGCCTACACACTGGCCTCCCTC 1489
     ||  :::||                     ||| |||
901  .............................MetGlyProProG 905
```

```
1490 TACCACCGGAGAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGG 1539
     ||||||| || |||::::||
905  lyProProGlyProLeuGlyIleProGlyLeu.....SerGly.ValPr 919

1540 AGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGGCCCTA 1589
     ||| :::::::::||::||:||:
919  oGlyLeuLysGlyAspAspGlyLeuGln...........GlyG 930

1590 AGCCTGGAGCTCCCTTCCCTAATGACACGTGGGTGCTGGAGGCAGTGC 1639
     ::||||| ||| |||:||:: |||:::: |||:::
930  lnProGlyLeuProGlyProThrGlyGluLysGlyGlySerLysGlyGluPro 946

1640 CTGCTCCCACCTCCACCCGCGTCTGCGGGGCCCTCTGCCTGTGATGCTC 1689
     ||| ||||||| |||:||
947  GlyLeuProGlyPro.Pro........GlyPro........MetAspP 957

1690 CGTACGTGTGGTGCGTGGGTGAGCCCACCGAGGCCAGGGTGGT........ 1731
     ||  ::: ||:::  ||||:::::||:||:
957  roAsnLeuLeu.GlySerLysGlyGluLysGlyGluProGlyLeuProGl 973

1732 ............TCCGGGCCG...GGGCAT.....CTGCCTGGA.....C 1756
            |||:::  ||:|| |||||||
973  yIleProGlyValSerGlyGlyTyrGlnGlyLeuProGlyAspP 990

1757 CTGCCCATCCTGCCGATAGTGC...CTTCCTGCTGTCCCAGGTGGCCCCATC 1803
     ||  ::||||||           |||:||||| ||||||
990  roGlyGlnProGlyLeuSerGlyLnProGlyLeuProGlyLeuProProGly 1006

1804 CCTGTTTATGGGCTCCATTGTCCAGCTCAGCCAGTCTGTCACTGCCTATA 1853
     ||| :::                  |||::|||||||
1007 ProLysGlyAsnProGlyLeuProGlyGlnPro.......... 1017

1854 TGGTGTCTGCGCCCAGGCCTGGGTCTGTCGCCATTTACTTGCTACACAG 1903
     |||::||                     |||
1018 .GlyLeuIleGlyProProGly..................LeuLysG 1027

1904 GTAGTATTTGCAAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTC 1953
     ||:::
1027 lyThr............................................. 1028

1954 CAGCACATTGGGGTGGAGGGCTGCGGGCTGGCCCGCCAGTTTCTTGTTGCTGCC 2003
     |||||||           ::: :::  |||:::|
1029 .....IleGlyAspMetGlyPhePro.GlyProGlnGlyIleGluGlyP 1043

2004 CTGTTAGCCCCATGGAGCTGCCGGCTGGCCGCCCAGTTTCTTGTTGCTGCC 2053
     ||::||||| ||||||:: ::||||||:::|
1043 roProGlyProSerGlyValProGlyLnProGlySerProGly.LeuPr 1059

2054 AAAGTAATGTGGCTCTCTGCTGCCACCCTGTGCTGCTGAGGTGCCTAGCT 2103
     |
1059 oGly............................................. 1060

2104 GCACAGCTGGGGCGTGGGGGCTGGCGTCCCTCCTCCTCCCCCAGTCTCTAGGG 2153
     ||   |||     ||||||::::||
1061 .GlnLysGlyAspLysGlyAspProGlyIleSer.....SerIleGly 1074

2154 CTGCCTGACTGGAGCCTTCCAAGGGG...GTTTCAGTCTGGACTTATAC 2200
     ||||||           |||||||          :::::
1075 LeuProGlyLeuProGlyProLysGlyGluProGlyLeuProGlyTyrPr 1091

2201 AGGGAGGCCA.........................GAAGGGCTCCATGCAC 2226
     |||:::|||                            :::|
1091 oGlyAsnProGlyIleLeuLysGlyGlySerValGlyAspProGlyLeuProGlyL 1108

2227 TGGAATGCGGGACTCTGCCAGGTGGATTACCCAGGGCTCAGGGTTAACAGC 2276
     ||      |||||||                   |||||||
1108 euProGlyThrProGlyAlaLysGlyGlnProGlyLeu............. 1120

2277 TAGCCTCCTAGTTGAGACACACCTAGAGAAGGGTTTTGGGAGCTGAATA 2326
     ||||||  |||  |||
1121 .........................ProGlyPheProGly........ 1125
```

```
2327 AACTCAGTCACCTGGTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCG 2376
         ::|||||||  |||:::   :::
1126 .......ThrProGlyProProGlyProLys............         1133

2377 TTTAATGTAGCTCTTGCATGGGAGTTTCTAGGATGAAACACTCCTCCATG 2426
         |||:::||| |||   |||||
1134 ..............GlyIleSer.Gly.........ProProGl 1140

2427 GGATTGAACATATGACTTATTGTAGGGGAAGAGTCCTGAGGGGCAACA 2476
      ::|
1140 yAsn.................................ProGlyLeuP 1145

2477 CACAGAAACCAGGTCCCCTCAGCCACCCAGCACTGTCTTTTTGCTGATCCA 2526
         |||||||||||||:::::::
1145 roGlyGluProGlyProValGlyLyGlyGlyHisProGlyGlnProGly 1161

2527 CCCCCCTTACCTTTTATCAGGATGTGGCCTGTTGGTCCTTCTGTTGCC 2576
      |||||
1162 ProPro...............................         1163

2577 ATCACAGAGACACAGGCATTTAAATATTAACTTATTATTTATTTAACAAAGT 2626
1163 ...................................         1163

2627 AGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGCTCTAATATTTGGGT 2676
                   |||
1164 ............................Gly 1164

2677 AGGGTGGGGGATCCCAACAATCAGGTCCCCTGAGATAGTGGTCATTGG 2726
      :::    ||| |||  |||:::|||
1165 GluLysGlyLysProGlyGlnAspGly.......         1173

2727 GCTGATCATTGCCA...........GAATCTTCTCTCCTGGGGTCTGG 2764
      ||| |||            :::|:::||||:::
1174 .......Ile.ProGlyProAlaGlyGlnLysGlyGluProGly...... 1185

2765 CCCCCAAAATGCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAAT 2814
      |||  :::   |||||||||||
1186 GlnProGlyPheGlyAsnProGlyPro..............         1194

2815 GATAATTCCAAAATGCTGTTACCCAAGGTTAGGTGTTGAAGGAAGGTAGA 2864
                        ||||::   |||  :::|||
1195 ..............Pro.........GlyLeuProGly..LeuSe 1201

2865 GGGT.......GGGGCTTCAGGTCTCAACGGCTTCCCTAACCACCCCCTCTT 2908
      :|||   ::::||||||   |||:::||||:::||::::
1201 rGlyGlnLysGlyAspGlyLeuProGlyIleProGlyAsnProGlyL 1218

2909 CTCTTGGCCCAGCCTGGTTCCCCCACTTCCAC.............. 2941
      |||||   |||:::||| |||||||
1218 euProGlyPro.LysGlyGluProGlyPheHisGlyPheProGlyValGl 1234

2942 .TCCCCTCTACTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAAT 2990
      :::|||
1234 nGlyPro.............................         1236

2991 TTCCCCTACCCCAACTTCCCCTACCCCCAACTTCCCCACCAGCTCCA 3040
      |||   |||||||   ||| :::|||
1237 ..ProGlyProProGlySerProGlyProAlaLeuGluGlyProLysGly 1252

3041 CAACCC 3046
      :::|||
1253 AsnPro 1254
```

seq_name: SwissProt_37:Ca11_BOVIN

```
seq_documentation_block:
ID   CA11_BOVIN     STANDARD;     PRT;   779 AA.
AC   P02453;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   COLLAGEN ALPHA 1(I) CHAIN (FRAGMENTS).
```

```
GN   COL1A1.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   ARTIODACTYLA; RUMINANTIA; PECORA; BOVOIDEA; BOVIDAE; BOVINAE; BOS.
RN   [1]
RP   SEQUENCE OF 1-19.
RX   MEDLINE; 72255334.
RA   RAUTERBERG J., TIMPL R., FURTHMAYR H.;
RT   "Structural characterization of N-terminal antigenic determinants in
RT   calf and human collagen.";
RL   EUR. J. BIOCHEM. 27:231-237(1972).
RN   [2]
RP   SEQUENCE OF 20-145.
RX   MEDLINE; 76022320.
RA   FIETZEK P.P., KUEHN K.;
RT   "The covalent structure of collagen: amino-acid sequence of the
RT   cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5
RT   from calf-skin collagen.";
RL   EUR. J. BIOCHEM. 52:77-82(1975).
RN   [3]
RP   SEQUENCE OF 146-294.
RX   MEDLINE; 73049499.
RA   FIETZEK P.P., WENDT P., KELL I., KUEHN K.;
RT   "The covalent structure of collagen: amino acid sequence of alpha-1-
RT   CB3 from calf skin collagen.";
RL   FEBS LETT. 26:74-76(1972).
RN   [4]
RP   SEQUENCE OF 295-562.
RX   MEDLINE; 74086118.
RA   FIETZEK P.P., REXRODT F.W., HOPPER K.E., KUEHN K.;
RT   "alpha-1-CB7 from calf-skin collagen. 2. The amino-acid sequence of
RT   alpha-1-CB7 from calf-skin collagen.";
RL   EUR. J. BIOCHEM. 38:396-400(1973).
RN   [5]
RP   SEQUENCE OF 563-675.
RX   MEDLINE; 73042276.
RA   WENDT P., MARK K.V.D., REXRODT F., KUEHN K.;
RT   "The covalent structure of collagen. The amino-acid sequence of the
RT   112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-
RT   skin collagen.";
RL   EUR. J. BIOCHEM. 30:169-183(1972).
RN   [6]
RP   SEQUENCE OF 676-751.
RX   MEDLINE; 73042275.
RA   FIETZEK P.P., REXRODT F.W., WENDT P., STARK M., KUEHN K.;
RT   "The covalent structure of collagen. Amino-acid sequence of peptide
RT   alpha-1-CB6-C2.";
RL   EUR. J. BIOCHEM. 30:163-168(1972).
CC   -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC       (FIBRILLAR FORMING COLLAGEN).
CC   -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC   -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC       BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC       HYDROXYAPATITE.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- THE COMPLETE CHAIN CONTAINS 1052 RESIDUES.
DR   PIR; A91193; CGBO1S.
DR   PROSITE; PS01208; VWFC; PARTIAL.
KW   EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW   GLYCOPROTEIN; COLLAGEN.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       9      9       CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT                                INVOLVED IN CROSS-LINKING.
FT   MOD_RES     103    103       HYDROXYLATION.
FT   CARBOHYD    103    103       GLUCOSYLGALACTOSE.
FT   MOD_RES     115    115       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     124    124       HYDROXYLATION (POTENTIAL).
FT   NON_CONS    145    146
FT   MOD_RES     274    274       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     346    346       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     424    424       HYDROXYLATION (POTENTIAL).
FT   MOD_RES     496    496       HYDROXYLATION (POTENTIAL).
```

```
FT   MOD_RES       658    658        HYDROXYLATION (POTENTIAL).
FT   MOD_RES       670    670        HYDROXYLATION (POTENTIAL).
FT   MOD_RES       726    726        HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
FT                                   ONLY HYDROXYLATED PRO IN POSITION X (IN
FT                                   THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
SQ   SEQUENCE      779 AA;  70346 MW;  7B1A9A69 CRC32;
```

```
alignment_scores:
       Quality:   321.00           Length:    858
         Ratio:    0.928             Gaps:     53
Percent Similarity:  40.326    Percent Identity:  26.690

alignment_block:
US-09-030-606-110/rev x CAl1_BOVIN

Align seg 1/1  to: CAl1_BOVIN  from: 1  to: 779
```

```
2365 GTTAAGGGGCTTAGA.................GATGGGAAACCAGGTGACTG 2331
     ::|||||| |||             |||||| ||||||
 102 MetLysGlyHisArgGlyPheSerGlyLeuAspGlyAlaLysGlyAsp.. 117

2330 AGTTTATTCAGCTCCCAAAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAG 2281
     ::||||| :||||||:                 ||||||
 118 ......AlaGlyProAlaGlyProLys................. 124

2280 GCTAGCTGTTAACCCTGGCCCTTGGGTAATCCA......CCTGCCAGAGTCC 2237
     |||:|||| |||:||:|||:                 |||||||::|||
 125 .........GlyGluProGlySerProGlyGluAsn 133

2236 CCGCAATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAGAC 2187
     :||   |||||||||
 134 GlyAlaProGlyGlnMetGlyProArgGlyLeuPro.............. 145

2186 TGAAACCCCTTGGAAGGCTCCAGTCAGGCAGCCAGGCCTAGAGACTGGGGAG 2137
     |||                  ::|||:|||  |||:||:||||||||
 146 .GlyPheProGlyProLysGlyAlaAlaGlyAlaGlyGluProGlyLysAlaGlyG 162

2136 AGAGGAAGGGGACGCGCCCAGCCCCCA...GCTGTGCAGTCACGCACCTCA 2090
     ||  ||||||::|||  ||||||  ||||||
 162 lu...ArgGlyValProGlyProGlyAlaValGlyProAlaGlyLys 177

2089 GCAGCACAGGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAA 2040
     :::|::||||:|||::|||   |||
 178 AspGlyGluAlaGlyAlaGlnGlyPro................. 186

2039 CTGCGGCGCCAGCCCCGCAGCCCCATGGGCGTAACAGGAGCGGGGAGCTGG 1990
     |||     |||:::|   |||:::|||:  ::|||||||       |
 187 ......ProGlyProAlaGlyProAlaGlyGlyAluArgG 197

1989 GA.................................CCCAGTGAG 1979
     ||                             |||:||
 197 lyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyPro 213

1978 GCAGGCCCTCCACCCAATGTGCTGGAAGTTTTCTACGCTGAGTATTTGG 1929
     |||||||||||| ::: |||||:  |||:::||
 214 AlaGlyProProGlyGlu..AlaGlyLys................. 222

1928 CCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCCACC 1879
     |||||| ||||:||||||:              ||
 223 ..............GlyProSerGlyAlaAr 241

1878 AGAC...CCAGGCCTGCGGCCAGACACCATATAGGCAGTGACAGACTGGC 1833
     ||  !|||||        ::|||   |||||||
 226 lnGlyValProGlyAspLeuGlyAlaPro.............. 235

1832 TGAGCTGACAATGGAGCCCCATAAACAGGATGGGGCCACCTGGGACAGC 1783
     ||||| ||||||           |||:::
 236 ..............GlyProSerGlyAlaAr 241

1782 AGGA...AGGCACTACTCCAGGATGGCGAGGTCGCAGGCAGATGCCCCGGCC 1736
     |||  |||:||||| |||||||             |||||| ||||||
 241 gGlyGluArgGlyPheProGlyLeuArgGlyValGlyGlyProGlyP 258
```

```
1735 CGGAACCACCCTG.GCCTCGGTGGCTCACCACCACCACCACC......ACACGT 1693
     ||            |||:||:||||                     ::|||
 258 roAlaGlyProargGlyAlaAsnGlyAlaProGlyAsnAspGlyAlaLys 274

1692 ACGGAGACATCACAGGACGCAGAGGCCCCGCAGAGCGCGGTGAGGTGGGAG 1643
     ::||||||       ||||||||||| |||  :||||   :||
 275 GlyAspAlaGlyValAlaProGlyAlaProGlySerGlnGlyAlaProGlyLe 291

1642 CAGGCCACTGCCT................CCAGCACCCACGT 1617
     :::  ::!|||                    |||:|||
 291 uGlnGlyMetProGlyGluArgGlyValAlaAlaAlaGlyLeuProGlyProLysG 308

1616 GTCCATTAGGG..........AAGGGAGCTCCAGGCTTAGGGCCTGGC 1579
     |||                  |||:|||  ::|||||
 308 lyAspArgGlyAspAlaGlyProLysGlyAlaAspGly..AlaProGly 323

1578 AGGAAGCTGGTCATCAGGCGTGTCCCACTGCTAGCACCTCCAGTGTCCCC 1529
     :::    |||       ||| |||:::||||:::!||||
 324 LysAspGlyValArgGlyLeuThrGlyProIleGlyProPro.GlyProA 340

1528 TCGGTATT.................................TGGGCAGGA 1512
     |||||:                                    ||||||
 340 laGlyAlaProGlyAspLysGlyGluAlaGlyProSerGlyProAlaGly 356

1511 ACACCTGCTTCTCCCGTGTGTAGAGGGAGGCCAGTGTGTAGGGCAGGATC 1462
     ||| :::|:||||||||:::|:|||              |||:::
 357 ThrArgGlyAlaProGlyAspArgGlyGluPro.........GlyPr 369

1461 TGCAGGCTGAGAGGTGAACCCGTGAGGCGGCTGAAGCTGTCACCAC 1412
     |||
 369 oProGly........................................ 371

1411 GGGCCACACTGTGGGACAGGCATGTGGCACCGGCAGCCACAGGAAAGCTG 1362
     |||                      |||  |||
 372 ..........................ProAlaGly.......... 374

1361 CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCCAGCCGG 1312
     ||||||  ::          :::||||||| |||||
 375 ..............................PheAlaGlyProProGl 380

1311 TCCATGACCCAGAGAGAAGAACCAGGAGATGGCCACTGCAGGAACAGCCC 1262
     ||:||    :: :::|:||||                        |||
 380 Y......AlaAspGlyGlnProGly.......................AlaL 388

1261 CAGGCTGCCCATCCGAACGCCCTTCATCATAGTGTCTCCGGCTCCGGTGC 1212
     |||    |||:::|       |||:::::|||              :::||||||
 388 ysGlyGluProGlyAspAlaGlyAlaLysGlyAspAlaLysGlyProProGly 404

1211 CCGGCTCAGCTCTGGGCACGCCCTGGTACA..............GCCCC 1177
     ||||||      ::: ||||||                      ||||||
 405 ProAlaGlyProProGlyProGlyProIleGlyAsnValGlyAlaPr 421

1176 TCGCCCACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCCA 1127
     ||||||   ||| |||            |||            :
 421 oGlyPro.........LysGlyAlaArgGlySerAlaGlyProProG 434

1126 GCTGCCACGTCAGCCACGAAGAGCCCGCGAAGGGTGCGGGGCATGCGGC 1077
     ::::|||:::  |||                        |||||:::|||
 434 lyAlaThrGlyPhePro...................GlyAlaAlaGly 443

1076 AGCACAGCT......GGTGCAGCCGGGGAAGCAGGAGCGCCCAGGTTCCGG 1033
     |||         ||| ::||||::|||||||  ||||||
 444 ArgValGlyProProGlyProSerGlyAsnAlaGlyProProGly...... 458

1032 AAAGCCAAGCCGGGCCCCGCATGGACAGCAGTGGGGCGCAAGGAGGGGGC 983
     |||||||||||||||| :::                     |||||:::
 459 ...ProProGlyProAlaGlyLysGlyGluGlyLySGlyProArgGlyG 474

982 CGACAGGCCCTTCTGCTGCTCGGTGTGGGGGCCCAGCCTGCCTCCTCAGCCA 933
    |||:::       ::||||||| ||||||        :::  ||||||
 474 luThrGly...ProAlaGlyArgProGlyLeuValGlyProGlyProGlyP 489
```

```
932  CCAGCAGTGCTGGCTGCTACGCAGTGAGCAAGAGATGAGGGTGAGCAGGCCA   883
     |||                         ||  ||||||| ::  |||
490  Pro.......................... GlyProAlaGlyGl          495

882  AAGAGGCACTCCT............................CCTGGGTGCCCAG   857
     ::|||:::||||                            |||||  |||
495  uLysGlyAlaProGlyAlaAspGlyAlaProGlyAlaProGlyThrProG      512

856  GT..........................AGGGGG........CCAGGGCAC      840
     ||                          ||||||        ||||||
512  lyProGlnGlyIleAlaGlyGlnArgGlyValValGlyLeuProGly...     527

839  TGGTGTCCCAGTCAATGCAGCAGGA.....GGTAGCCCAGGCAGCCC          796
     |||     ::: |||::|||     |||  |||  ||||||   |
528  ..............GlnArgGlyGluArgGlyPheProGlyLeuPr          538

795  CCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCACAGTGGTC        746
     |                     :::::||||||| :::|||
538  oGlyPro.................SerGlyGluProGlyLysGlnGlyP       549

745  CGGGTCCCGGAAGAGGTCAGAGAGCCAGGCCCTCCAGTGAGTGAAGCACA       696
     ||::  ::: :||::|||  ||||||||          ::
549  roSerGlyAlaSerGlyGluArg..GlyProProGlyProMetGlyPro      564

695  CCTGGCCACAGAAGTCCA......GCAGCCCCACGCCCAGGATGAGCAGT      652
     ||||                    ||::|||       ||| |||
565  ProGlyLeuAlaGlyProProGlyGluSerGlyArgGluGly.......       578

651  GCCAGCTCCAGGGGCC...TGGGATCCGGGCACAGCAGCCCTGCTAGCCA      605
     ||||||||||||     ::::|||||            :::
579  ...AlaProGlyAlaGlyGluGlySerProGlyArgAspGlySerProGlyA    594

604  GCCGGCCCTTGGGATGAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTG      555
     ::|:::           |||  ||||||| ::::|||
594  laLysGlyAspArgGlyGluThrGlyGlnGlyProAlaGlyProProPro     610

554  CCCAGATGAAGGGCCGGCGGCGGCGGCCATAGCGTCCACGCCAGTGGTCACTG   505
     |||||||||||                            |||
610  ...................................................   610

504  GCTCGAGCCTAGGAGGCGGGACACAGACCCAGGCCCAGCCACTGGACCAATGCC   455
     ||||||            ||||||||:::
611  .........GlyAlaProGlyAlaProGlyProVal...........GlyP   621

454  CAGCACCATGGTCA.............TGAACTTCTCCTCTACCCCCACTT   417
     |||||               |||||||||          |||:::|||
621  roAlaGlyLysSerGlyAspArgGlyGluThrGlyAspGlyProAlaGlyProIle   637

416  ...CCAGCCAGAGCGGCCGGCACATAGGTAGTCCTGCGGCCAAACACACC   370
     ::|:::           ||||| ||||||              |||  ||
638  GlyProValGlyProAlaGlyAlaArgGlyProAlaGlyProGlnGlyPr     654

369  TCCAGGCCAAGG.........................TTAGCAG         350
     ||::::::::                           :::
654  oArgGly***LysGly***ThrGly******Gly***ArgGlyIleLysG    671

349  GTTGACCAGCAAGAGCTGGCTTTCCGGT...GCCGCAGCAGGCGGCTCA      303
     |||||||        ::|||||||||:::|||      |||||
671  lyHisArgGlyPheSerGlyLeuGlnGlyProProGlyProGlySer        687

302  CCCACAGCCTCTGGACCATAGTGGG........................   278
     |||    ::|||||| ||||||
688  ProGlyGluGlnGlyPro.SerGlyAlaSerGlyProAlaGlyProArgG     704

277  .....CCAGGCGGGTAGGCTCAGGGGGCGTTCAGGCACTCCAGAACTG        233
     |||||||::  ||||||   ||:::::::|||
704  lyProProGlySerAlaGlySerProGlyLysAspGly...........      716

232  CTTCGTCTCGGCCTCTGCTCCAGAAGCTGCGGCCTCTCCTCCTTGCTGCCG   183
     |||:::::: |||    |||    :::  |||  |||
717  .....LeuAsnGlyLeuProGlyProIleGlyProProGly........     728

182  CCAACTGCCTAGGAATCAGCCAGCGCCCATTCTGCCAGCCCTTTGGTG     133
```

```
729  .......ProArgGlyArgThrGly.......................... 729
         |||||||:::::: |||
         ...ProArgGlyArgThrGly.......................AspA 736

132  CCGGTCCAGCTTCTCAGCCC  113
     ||||||||||||:::   ||
736  laGlyProAlaGlyProPro  742
```

OM of: US-09-030-606-110 to: SPTREMBL_10:*   out_format : pfs

Date: Sep 25, 1999 10:32 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09030606/runat_24091999_171617_29869/app_query.fasta.1
-DB=SPTREMBL_10 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.000 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-110
Query length: 3410
Database: SPTREMBL_10:*
Database sequences: 201082
Database length: 61543640
Search time (sec): 260.540000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
sp_virus:O41971		382.50	489.33	2.9e-20	774	! O41971 murine herpesvirus 68. h
sp_virus:O41972		365.50	467.73	5.1e-19	706	! O41972 murine herpesvirus 68. h
sp_invertebrate:O26634	-	364.00	459.88	7.0e-19	1414	! Q26634 strongylocentrotus pu
sp_virus:O41973		359.00	458.92	1.5e-18	727	! O41973 murine herpesvirus 68. h
sp_rodent:Q63079	-	357.00	450.43	2.3e-18	1453	! Q63079 rattus norvegicus (rat)
sp_vertebrate:O93251	-	354.50	447.19	3.5e-18	1445	! O93251 rana catesbeiana (bull
sp_plant:Q65803	+	347.50	446.70	1.0e-17	515	! Q65803 daucus carota (carrot).
sp_human:Q76045	-	342.50	431.29	2.7e-17	1461	! Q76045 homo sapiens (human).
sp_plant:O04077	+	337.50	433.40	5.7e-17	523	! O04077 vicia faba (broad bean).
sp_vertebrate:Q92029	-	335.50	426.31	8.3e-17	886	! Q92029 gallus gallus (chicken).
sp_rodent:Q60444	+	335.50	421.58	8.7e-17	1549	! Q60444 cricetulus griseus (chi
sp_rodent:Q88207		335.50	420.14	8.8e-17	1838	! Q88207 mus musculus (mouse). c
sp_rodent:Q60467	-	334.50	418.81	1.0e-16	1840	! Q60467 cricetulus longicaudatu
sp_plant:Q9ZTB9	+	333.00	427.66	1.2e-16	512	! Q9ztb9 apium graveolens (celery
sp_plant:O43653	+	333.00	427.59	1.2e-16	516	! O43653 solanum tuberosum (potat
sp_invertebrate:O97405	-	333.00	419.28	1.3e-16	1378	! O97405 haliotis discus (abal
sp_human:Q15094		333.00	416.84	1.3e-16	1838	! Q15094 homo sapiens (human).
sp_plant:O80605	+	328.50	420.47	2.6e-16	594	! O80605 arabidopsis thaliana (mo
sp_mammal:Q28396		328.00	412.46	3.1e-16	1418	! Q28396 equus caballus (horse).
sp_plant:Q41152	-	326.00	418.10	4.0e-16	533	! Q41152 ricinus communis (castor
sp_human:Q14047		325.50	408.76	4.7e-16	1487	! Q14047 homo sapiens (human). a
sp_vertebrate:Q9YIB4	-	325.00	408.32	5.1e-16	1450	! Q9yib4 cynops pyrrhogaster (ja
sp_vertebrate:O42350	-	324.50	408.23	5.5e-16	1355	! O42350 rana catesbeiana (bull
sp_rodent:O39231	-	323.50	415.15	6.0e-16	512	! O39231 arabidopsis thaliana (mc
sp_plant:O80550	+	323.50	415.15	6.0e-16	512	! O80550 arabidopsis thaliana (mc
sp_human:Q14046	+	323.00	407.57	7.0e-16	1160	! Q14046 homo sapiens (human). c
sp_rodent:Q61431		323.00	405.41	7.0e-16	1497	! Q61431 mus musculus (mouse).
sp_mammal:O46392		321.00	403.55	1.0e-15	1366	! O46392 canis familiaris (dog).
sp_invertebrate:O97406	-	321.00	403.11	1.0e-15	1439	! O97406 haliotis discus (abal
sp_plant:O40583	-	319.50	409.96	1.2e-15	507	! O40583 nicotiana tabacum (commo
sp_rodent:Q62031	-	320.00	401.78	1.2e-15	1442	! Q62031 mus musculus (mouse). E
sp_rodent:Q62032	-	316.00	396.41	2.3e-15	1459	! Q62032 mus musculus (mouse). E
sp_rodent:Q63123	-	315.50	395.99	2.5e-15	1419	! Q63123 rattus norvegicus (rat)
sp_rodent:Q62033	-	312.00	391.25	4.6e-15	1442	! Q62033 mus musculus (mouse). E
sp_rodent:Q61431	+	311.50	390.27	5.0e-15	1497	! Q61431 mus musculus (mouse). E
sp_rodent:Q63870	-	311.00	383.89	5.8e-15	2944	! Q63870 mus musculus (mouse). E
sp_invertebrate:Q17163	+	310.50	387.39	6.0e-15	1802	! Q17163 brugia malayi. a2 (iv
sp_invertebrate:O16161	-	309.50	391.92	6.7e-15	902	! O16161 mytilus edulis (blue m
sp_plant:Q65929	-	309.00	396.24	7.6e-15	501	! Q65929 daucus carota (carrot).
sp_mammal:O62649		309.00	387.77	8.0e-15	1364	! O62649 bos taurus (bovine). al
sp_vertebrate:Q91717		308.00	385.73	9.0e-15	1486	! Q91717 xenopus laevis (african
sp_rodent:Q9Z1W0	-	307.00	383.52	1.1e-14	1650	! Q9z1w0 mus musculus (mouse). c

sp_invertebrate:Q07265	-	307.00	306.50	383.02	1.1e-14	1752	! Q07265 strongylocentrotus
sp_vertebrate:Q91718	-		304.50	383.72	1.2e-14	1491	! Q91718 xenopus laevis (afri
sp_human:Q14054	-		364.00	375.43	1.7e-14	2912	! Q14054 homo sapiens (human)

seq_name: sp_virus:O41971

seq_documentation_block:
ID O41971   PRELIMINARY;   PRT;   774 AA.
AC O41971;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 73.2 KD PROTEIN.
GN GAMMAHV.M10A.
OS murine herpesvirus 68.
OC viruses; dsDNA viruses, no RNA stage: Herpesviridae;
OC Gammaherpesvirinae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WUMS;
RX MEDLINE; 97366649.
RA VIRGIN H.W. IV., LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E.,
RA DAL CANTO A.J., SPECK S.H.;
RT "Complete sequence and genomic analysis of murine gammaherpesvirus
RT 68.";
RL J. Virol. 71:5894-5904(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WUMS;
RA LATREILLE P., WAMSLEY P., WATERSTON R.H.;
RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U97553; AAB66420.1; -.
KW Hypothetical protein.
SQ SEQUENCE 774 AA; 73163 MW; 6A851113 CRC32;

alignment_scores:
                Quality: 382.50   Length: 1097
                Ratio: 0.935   Gaps: 74
  Percent Similarity: 37.284   Percent Identity: 25.980

alignment_block:
US-09-030-606-110/rev x O41971

Align seg 1/1 to: O41971 from: 1 to: 774

2939 GGAAGTGGGGGAACCAGGCTGGCCAAGAGAGAGAGGGTGGTTAGGGAA 2890
     ||::|||||| :::|||||||||::::|||||| |||
  31 GlyThrGlyGly......ThrGlyProArgSerGluGly...AlaGlyCy 44

2889 GCCGTTGAGACCTGAAGCCCACCCTCTACCTTCCTTCAACACCCTAACC 2840
     |||:|||||| ::|||||||:::||||::
  44 sProAlaArgGlyAlaGlyProProAla.......................... 53

2839 TTGGGTAACAGCATTGGAATTATCATTGGGATGAGTAGAATTCCAAG 2790
     53 53

2789 GTCCTGGGTTAGGCATTTGGGGGCCAGACCCCAGGAGAAGAAGATTCT 2740
     ||||||||| ||| ::::||
  54 ....TrpGlyAlaGlyProProArgArg.......... 61

2739 GGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGACCTGATTGTTGGG 2690
  61 61

2689 GATCCCCCACCTACCCAAATATTAGACACCACCAGAAAGCTAGCAA 2640
  61 61

2639 TGGATTCCTTCTACTTTGTTAAATAATAAGTTAAATATTTAAATGCCT 2590
  61 61

```
2589 GTGTCTCTGTGATGGCAACAGAGGACCAACAGGCCACATCCTGATAAAA 2540          1782 AGGAAGGCACTATCCAGGATGGCGAGGTCAGGCAGATGCC............ 1742
            |||||||:::::|||                                                  ||| ||:::||         :::           |||||||
  62 ...........AspGlyGlyAsnArg...............          66            262 pGly.GlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeuPro 278

2539 GGTAAGAGGGGGTGGATCAGCAAAAGACAGTGCTGTGGCTGAGGGGA 2490          1741 ...CCGGCCGGACCACCCTGGCCTCGGTGGGCTCACCCGTCACCCACCAC 1695
            |||  :::||| :::::|||   |||||| |||                                   ::: ||| |||||||||:::     :::
  67 ...........GlyAspGlyAlaProGluArgGluArgGlySerGlyLeuProGly 79        279 GlyProGlyCysGlyLeuGlySerArgSerProAlaGly 295

2489 CCT......GGTTCTTGTGTGTTGCCCCTCAGGACTCTTCCCCTACAAAT 2446          1694 GTACGGAGACATCAGAGCAGGAGGCCCCCCAGAGCGC............... 1658
     |||      ||||| |||   ||| ||||||::: |||                            :::||:::||   |||||||||::  |||
  80 ProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGly.. 95            295 yTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgAlaAlaAla 312

2445 AAGTCATATGTTCAAATCCCATGGGAGGAGTGTTTCATCCTTAGAAACTCCC 2396          1657 .....GGTGGAGGTGGGAGCAGGCAGGCCACTGCCTCCAGCCACCCAGTGTCC 1613
     :::::        ||||||       |||:::                                 ||| ||| ||||||     |||::::|||
  96 ...........TrpGlyGluProGlyGly..........ArgGlyProG 105            312 rgProGlyValArg.ValProArgProGlyProValProGlyLeuProGlyG 328

2395 ATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGGGCTTAGAGATGG 2346          1612 ATTAGGGAAGGGAGCTCCAGGCTTAGGGCCTGGCCAGGAAGCTGGTCATCA 1563
         ||||||        ||| ||:::|||                                      ::::|||           ||| |||  :::::
 105 lyAlaArg.......GluArgAlaAlaArg.ProGlyValArgValPr 118            328 yMetGly..GlyThrGlyGlyThrGlyPro..Arg.SerGluGlyAla 342

2345 GAAACCAGGTGACTGAGTTTATTCAGCTCCCAAAAAACCCTTCTCTAGGTG 2296          1562 GGCTGTCCTCACTGCTAGCACCTCCAGTGTCCCCTCGGTATTTGGGCAGG 1513
     ::||||||:::|||    |||                                            |||||||||          |||   ||| |||||
 118 oArgProGlyGluProVal..............ProProGlyGlyM 129            343 GlyCysProAla......ArgGlyAlaGlyProProAlaTrpGlyAlaGl 357

2295 TGTCTCAACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATCCACCT 2246          1512 AACACCTGCTTCTCCCGGTGGT..AGAGGGAGGCCAGTGTGTAGGGCAG 1466
     :::::         |||||| |||                                         || |||           ||| :::  |||||
 129 etGlyGlyThrGlyGly..........          134            357 yProProArgArgAspGlyGlyAsnArgGlyAspGlyAlaProGluArgG 374

2245 GCAGAGTCCCCGCATTCCAGTGCATGGAAGGCCTCCTGGCCTCCCTGTATA 2196          1465 GATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCCGGCTGAAGCTGTCA 1416
         |||  :::|||:::|||:::::|||   |||:::::|||                           ||||||   |||||||
 135 ...ThrGlyProArgSerGluGly.......AlaGlyCysProAla.A 147            374 lySer...GlyLeu...........          377

2195 AGTCCAGACTGAAACCCCCTTGGAAGGCCTCCAGTGCAGGCAGCCCTAGAG 2146          1415 CCACGGCCACACTGTGGGACA...GGCATGTGGCACCGGCAGCCACCAGGGG 1369
     :::::    ||| ||| |||||| ||| :::                                  ||| ||| |||||||:::  :::|||
 147 rgGlyAlaGlyProProAlaTrpGlyAlaGlyProProAlaGlyPro..Asp 162          378 ProGlyProGlyCysGlySerProGlyLeuGlySerArgSerPro..... 392

2145 ACTGGGGAGGAGAGGAGAGGGGACGCC................CCAG 2117          1368 AAAGCTGCCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGCTCCAC 1319
     ||||||:::|||||| ::||| |||                                         |||||||||
 163 GlyGlyAsnArgGlyAspGly.AlaProGluArgGlySerGlyLeuProG 179          393 ................................P 393

2116 CCCCCAGCTGCAGGCTACGCACCTCAGCAACAGAGGTGGCAGCAGAGA 2067          1318 CAGCCGGTCCATGACCAGGAAGAAGAGAGACCAGGGAGATGG...CGCACTGCA 1272
     ::||||::::|| |||||::||||||                                         |||||||         :::    |||||||||
 179 lyProGlyCysGlySerProGlyLeuGlySer.......          189            393 roAlaGly.................TrpGlyGluProGly 400

2066 GCCACATTACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCAGCCCC 2017          1271 GGAACAGCCCCAGGCTGCCCATCCGAACGCCTTCATCATAGTGTCTCCGG 1222
                                        ||||||||                               ||| :::|||||||:::    ||||||
 190 ........................ArgSerProProAlaGl 195            401 GlyArgGlyProGlyAla..ArgGluArgAla........... 410

2016 ATGG.......................GGCTAACAGGAGGAGGGGGGAGCT 1992          1221 GCCTCCGGTGCCCGGCTTCGGGCACGCCCTGGTACA...GCCCCTC 1175
          |||                          ||| :::||||||:::||                              ||||||||||||   :::|||
 195 yTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgAlaAlaAlaA 212          411 ..................AlaAlaProGlyValArgValProA 419

1991 ....GGGACCCAGTGAGCGAGGCAGGCCCTCCACCCCAATGTGCTGGAAGTTT 1947          1174 GCCCCACGAAATCCGTGTAAAAACAGCTGAAGGTCATGAGTGCCA....... 1131
     ||:::::||         |||||||:::|||  :::|:|                            ||||| :::||||||||||||||||||
 212 rgProGlyValArgValProArgProGlyGluProVal.......          224            419 rgAspGlyGlyAsnArgGlyAspGlyGlyProGlyAlaProGluArgGlyLeu 477

1946 TCTACGCTGAGTATTTGGCCAAGTCGCCTGTTCGTCAAATACTACCTGTGTA 1897          1130 .........TCCAGCTGCACAGCTCAGCCACGAAGAGCCCGGCGCAGGGTG 1090
                                                                            |||||         |||||||||||||||
 224 ................................          224            428 GlyMetGlyGlyThrGlyThrGlyProArgSerGluGlyAlaGlyCy 444

1896 GCAAAGTAAATGGCACCAGACCAGACCCCAGGCCCTGCGGCAGACA........ 1857          1089 CGGGGCA.......TGCGGCAGCACAGCTGCTGTCAGCCCGGG.......... 1056
         |||            |||:::||                                         :::::|||         ||| :::||||||
 225 ................ProProGlyGlyMetGlyThrGlyGlyThrGlyThrG 236          444 sProAlaArgGlyAlaGlyProProAlaTrpGlyAlaGlyProProArgA 461

1856 ...................CCATATAGGCAGTGACAGACTGGC 1833          1055 ..GAAGCAGGGCGCCCAGGTTCCGGAAAGCCAAGC.............. 1023
                            ||| |||                                     :::::|||  |||:::|||   ||||:::
 236 lyProArgSerGluGlyAlaGlyCysProAlaArg..............          247            461 rgAspGlyGlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeu 477

1832 TGAGCTGGACAATGAGCCCATAAACAGGATGGGGCCACCTGGGACAGC 1783          1022 ...GGGCCCGGCATGGACAGCAGTGGGGCGACAAGGAGGGGCCCACAGC 976
         ||| ||||||                   || ||||||||                         |||||||          ||:::||| ||
 248 .....GlyAlaGlyProProAlaTrpGlyAlaGlyProProArgArgAs 262            478 ProGlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAl 494

                                                                        975 C..................................CTTCTGCTGG 965
```

```
 494  aGlyTyrGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgAlaA  511          709  rgAlaAlaAlaArgProGlyVal.......ArgValProGlyVal  721

 964  CTCGGTGGGGCCCAGCGCTGCCT.............CCTCAGCGACCAGCA  927          157  GCCCATTTCTGCCAGCCCCTTT...................GG   135
      |||||  ||||      :::|||             ||| ||||||::::                     :::|||
 511  laArgProGlyValArgValProArgProGlyGluProValProProGly  527          722  .........GluProValProProGlyMetGlyThrGlyGl  734

 926  GTGTGGCTGCTACGCCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGG  877          134  TGCCGGTCCAGCTTCTCAGCCCCATGCTCAACACCTGCTGC.........   95
                            ||| :::|||          ||               ::::|||||| ||:::            :::::|||
 528  ...........GlyMetGlyGlyThrGlyGl  534          734  yThrGlyProArgSerGlu........GlyAlaGlyCysProAlaArgG  748

 876  CACTCCTCCT............GGGTGCCCAGTAGGGGGCCAGGGC  842           94  .TGTGGGGCACCTCAGTGGGA.............CACGTCTCATCA   62
      ||| |||               ||||||||||:::|||||  ||:::                    ::::||| ||| |||||      |||||
 534  yThrGlyProArgSerGluGlyAlaGlyCysProAlaArgGlyAlaGly.  550          748  lyAlaGlyProProAlaTrpGlyAlaAlaLysAlaArgSerThrAlaGlu  764

 841  ACTGGTGTCCCAGTCAATGGCAGGCAGGA...................  813           61  CTCAGATCCTGGCCAGGCGCGCCGGCT  35
      |||   |||   ||||||||                                                 ::|||
 551  ........ProProAlaTrpGlyAlaGlyProProArgArgAspProGly  564          765  ValArgProValAlaArgArgGlyAla  773
```

```
 765  TAGGCCTTGGCACAGTGGTCCGGT.........CCCGGAAGAGGTCAGA  725          alignment_scores:
      ::|||                             |||||| ||| |:::                        Quality:  365.50          Length:   1065
 576  .......GlyLeuProGlyProGlyCysGlySerProGlyLeuGlySerA  590                    Ratio:   0.944          Gaps:       68
                                                                        Percent Similarity: 36.338   Percent Identity: 25.540
 724  GAGCAGGGCCCTTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCA  675
      ||:::  |||||||::   ::: |||||| ||:::||:::                  alignment_block:
 590  rgSer...ProProAlaGlyTrpGlyGluProGlyGlyArgGlyProGly  605          US-09-030-606-110/rev x O41972

 674  ...........GCCCCACGGCCCAGGATGAGCAGTGCCAGCTCCAGGG..  639          Align seg 1/1  to: O41972  from: 1   to: 706
                ||| |||||||| ||  ||||||  |||||||||
 606  AlaArgGluArgAlaAlaArgProGlyValArgValProArgProGlyGl  622          2939  GGAAGTGGGGGAACCAGGCTGGGCCAAGAGAGAGGGGTGGTTAGGGAA  2890
                                                                        ||||||||||||| :::||||||||||||||| :||||| |||||
 638  .GCCTGGGATCCGGCACAGCAGCCCTGCTAGCCAGCCGACCGGCCCTTGGGAT  590          3   GlyThrGlyGly......ThrGlyProArgSerGluGly...AlaGlyCy   16
      |||||
 622  uProValProProGly  627          2889  GCCGTTGAGACCTGAAGCCCCACCCTCCTTCAACACCCTAACC  2840
                                                                        ||| |||      ::|||||::|||||
 589  GAGAAAGAGGGCTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAAGGGCC  540          16  sProAlaArgGlyAlaGlyProProAla.............  25
                 ||| ||||||        ||||||||||
 628  .........GlyMetGlyGlyThrGlyGlyThrGlyProArg.........  638          2839  TTGGGTAACAGCATTGGAATTATCATTTGGGATGAGTAGAATTTCCAAG  2790
                                                                        |:::|||
 539  GGCGGCGGCCCATAGCCGTCCCACCGCCAGTGGTCACTGGCTGAGCCTAGGAGC  490          25  sProAlaArgGlyAlaGlyProProAla.......   25
      ||||||| |||||
 639  ...............SerGluGlyAl  642          2789  GTCCTGGGTTAGGCATTTGGGGGGCCCAGACCCCAGGAGAAGAAGATTCT  2740
                                                                        :::::|||||| |||
 489  GGGACACAGACAGGCCAGCACTGACCAATGCCCAGCACCACCATGGTCAT  440          25
      |||||||||
 642  aGly...........CysProAla..........  646
```

```
 439  GAACTTCTCCTCTACCCCCACTTCCAGCAGCAGAGAGGCGGCACATAGGTGA  390
      ||||||           :::|||
 647  ..ArgGlyAlaGlyProProAlaTrpGlyAlaGlyPro...........  658

 389  TGCCTGCGGCCCAAACACACCTCCAGGCCAAAGTTAGCAGGTTGACCAGC  340
      ::|:::     ||| |||     ::||||| |||||
 659  .........ProArgArgAspGlyAsnArgGlyProGlyAlaProGlI  672

 339  AAGA...GCTGGGCTTTCCGGTGCCCAGCAGGC......GGCTCACCCA  299
      ||:::  ||||||| |||  :::|||||||
 672  uArgGlySerGlyLeuProGlyCysGlySerProGlyLeuGlyLys  689

 298  CAGCCCTCGGACCATAGTGGG.........CCAGGGCGGTAGGGCTTCAGG  258
      :: |||::|||        ||||||||||      |||  ||||||
 689  erArgSerProPro.AlaGlyTrpGlyGluProGlyGlyArgGlyProGlI  705

 257  GGGGCCGGTTCAGGCACTCAGGAACTGCTTCGTCTCCGGCTCTGCTCCAGAAG  208
      |:::|||                        |||
 705  yAlaArg.............GluA  709

 207  CTGCGGCCTCCTCCTCCTTGCTGCGCGCCAACTGCCTAGGAATCAGCCAGGC  158
      |||||||       :::::|||||| |||
```

```
 26  ...................................................  33
                   |||||||| |||   :::::
                   ............TrpGlyAlaGlyProProArgArg.......  33
2739 GGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGACCTGATTGTTGGG  2690

 33  ...................................................  33
2689 GATCCCCACCCTACCCAAATATTAGACACCAACACAGAAAAGCTAGCAA   2640

 33  ...................................................  33
2639 TGGATTCCCTTCTACTTTGTTAAATAAATAAGTTAAATATTTAAATGCCT   2590

 33  ...................................................  33
2589 GTGTCTCTGATGGCAACAGAGAAGGACCAACAGGCCACATCCTGATAAAA   2540
                  ||||||:::::||
 34  .........................AspGlyGlyAsnArg............  38
2539 GGTAAGAGAGGGGGTGGATCAGCAAAAAGACAGTGCTGTGGGCTGAGGGGA  2490
                  |||  ::::||    :::::||       |||
 39  .............GlyAspGlyAlaProGluArgGlySerGlyLeuProGly  51
2489 CCT......GGTTCTTGTGTGTTGCCCCTCAGGACTCTTCCCCTACAAAT   2446
        |||||   |||                     :::||:::   |||
 52  ProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGly..   67
2445 AAGTCATATGTTCAAATCCCATGGAGGAGTGTTTCATCCTAGAAACTCCC   2396
        :::   ||||||||       |||:::||
 68  ..........TrpGlyGluProGlyGly..........ArgGlyProG    77
2395 ATGCAAGAGCTACATTAAACCAAGCTGCAGGTTAAGGGCTTAGAGATGG   2345
        ||||||       :::||||:||       ||:::||
 77  lyAlaArg........GluArgGlyAlaAlaArg.ProGlyValArgValPr  90
2345 GAAACCAGGTGACTGAGTTATTCAGCTCCCAAAAAACCCTTCTCTAGGTG  2296
        ||||||       ||                      ::||
 90  oArgProGlyGluProVal.......................ProProGlyGlyM 101
2295 TGTCTCAACTAGGAGGCTAGCTGTGTTAACCCTGAGCCTGGGTAATCCACCT 2246
        :::::  ||||||
101 etGlyGlyThrGlyGly.................................  106
2245 GCAGAGTCCCCGCATTCCAGTGCATGGAGCCCCTTCTGGCCTCCCTGTATA 2196
        :::||:::||:::       |||:::||  ||||:::: :
107 ...ThrGlyProArgSerGluGly.........AlaGlyCysProAla.A  119
2195 AGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGCAGCCCTAGAG  2146
        :::::  ||| ||| ||:::     ||||||:::
119 rgGlyAlaGlyProProAlaTrpGlyAlaGlyProProArgArg...Asp  134
2145 ACTGGGGAGGAGGAGAGGGGACGCGCC................CCAG    2117
        |||:::||||||:::||| |||                    ::
135 GlyGlyAsnArgGlyAspGly.AlaProGluArgGlySerGlyLeuProG  151
2116 CCCCCAGCTGTGCAGCTACGCACCTCAGCAGCACCAGGGTGGCAGCAGAGA 2067
        |||     :::||||||:::||||
151 lyProGlyCysGlySerProGlyLeuGlySer...................  161
2066 GCCACATTACTTTGGCAGCAACAGAAACTGGCGGCCCAGCCGGGCAGCCCC 2017
        |||||:::||||||:::::                   |||||   |||
162 .............................ArgSerProProAlaGl     167
2016 ATGG.................................GGCTAACAGGAGCGGGGAGCT 1992
        |||                             :::||||||:::||||:::||
167 yTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgGlyAlaAlaA 184
1991 ....GGGACCCAGTGAGGCAGGCCCTCCACCCCAATGCTGCTGGAAGTT 1947
        :::||||||||::::||       ||:::||::::
184 rgProGlyValArgValProArgProGlyGluProVal..........   196
1946 TCTACGCTGGAGTATTTGGCCAAGTCGCTCTTGTCAAATACTACCTGTGTA 1897
```

```
196  ...................................................  196
                   |||||||| |||   :::::
197  ...............................TrpGlyAlaGlyProProArgArg 208
1896 GCAAAGTAAATGGCGACCAGACCCAGGCTGCGGGCAGACA........... 1857
        |||   |||:::        ||||||
197  ...............................ProProGlyGlyMetGlyThrGlyThrG 208
1856 ....................................CCAATAGGCAGTGCAGTGACAGACTGGC 1833
                                          |||   |||
208  lyProArgSerGluGlyAlaGlyCysProAlaArg.................. 219
1832 TGAGCTGGACAATGGAGCCCATAAACAGGATGGGCCACCTGGACAGC    1783
        |||  |||                      |||
220  .....GlyAlaGlyProProAlaTrpGlyAlaGlyProProArgArgAs   234
1782 AGGAAGGCACTATCCAGGATGGCGAGGTCCAGGCAGATGCC.........  1742
        |||  |||:::                  :::
234  pGly.GlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeuPro  250
1741 ...CCGGCCCGGAACCACCCTGGGCTCGGTGGGCTCACCCACCACCAC    1695
        :::         |||||||:::|||       : : :
251  GlyProGlyCysGlySerProGlyLeuGlySerArgSerProProAlaGl  267
1694 GTACGGAGACATCACAGGCAGGAGGCCCCGCCAGAGCGC...........  1658
        :::|||:::       ||||||||||:::       | |
267  yTrpGlyGluProGlyGlyArgGlyProGlyAlaArgGluArgGlyAlaAlaA 284
1657 .....GGGTGGAGGTGGGACGCAGGCCACTGCCCTCCAGCACCCACGTGTCC 1613
        |||   |||       :::|||       :::||:::|||
284  rgProGlyValArg.ValProArgProGlyGluProValProProGlyGl  300
1612 ATTAGGGAGGGACGCTCCAGGCTTAGGGCCTGCCAGGAGCTGTGTCATCA  1563
        |||   |||       |||||:::||       : :  :::
300  yMetGly..GlyThrGlyThrGlyPro..Arg.SerGluGlyAla      314
1562 GGCTGCTCCTCACTGCTAGCACCTCGCAGTGTCCCCTCGGTATTGGGCAGG 1513
        ||||||||       |||::       ||||  :::
315  GlyCysProAla.....ArgGlyAlaGlyProProAlaTrpGlyAlaGl   329
1512 AACACCTGCTTCTCCGGTGGT...AGAGGGAGGCCAGTGTGTAGGGCAG   1466
        ::::||||       |||||||       :::  :::||||
329  yProProArgArgAspGlyGlyAsnArgGlyAspGlyAlaProGluArg   346
1465 GATCTGCAGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGTCA   1416
        |||||:::||||||  |||:::||:::|||:::|||
346  lySer..GlyLeu......................................  349
1415 CCACGGCCACACTGTGGGACA...GGCATGTGGCACCCGGCAGCCACAGGG 1369
        |||   |||  ||||||       :::  :::||::::||
350  ProGlyProGlyCysGlySerProGlyLeuGlyLySerArgSerPro...... 364
1368 AAAGCTGCCACACTGCCCAAATAGACTGCTCGACTGCCGAATCGCTGCAC  1319
365  ...................................................  P 365
1318 CAGCCGGTCCATGACCAGGAAGAAGAGAAGACCAGGGAGATGG...CGCACTGCA 1272
        |||||||                                 |||
365  roAlaGly.................................TrpGlyGluProGly 372
1271 GGAACAGCCCCAGGCTGCCATCCGAACGCCTTCATCATAGTGTCTCCGG  1222
        |||||   ::||||||:::::                     ||||||
373  GlyArgGlyProGlyAla..ArgGluArgGlyAla...............  382
1221 GCCTCGGTGCCCGGCTCAGCTCTGGCACGCCCTGGTACA...GCCCCTC  1175
        |||||||||||||:::       :::|||
383  .........................................AlaArgProGlyValArgValProA 391
1174 GCCCACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCA...... 1131
        |||||     :::|||         ::::||
391  rgProGlyGluPro..............ValProProGly           399
1130 ........TCCAGCTGCACAGCTCAGCCACGAAGAGCCCGGCCCAGGGTG  1090
        :::       ||||||       |||||:::||||||||||||
400  GlyMetGlyGlyThrGlyThrGlyProArgProGlyArgSerGluGlyAlaGlyCy 416
```

```
1089 CGGGCA......TGCGGCAGCACAGCTGTGCAGCCGGG........... 1056
      |  |  ::||||   |||   ::|||||
416  sProAlaArgGlyAlaGlyProProAlaTrpGlyAlaGlyAlaGlyProProArga 433

1055 ..GAAGCAGGGCGCCCAGGTTCCGAAAGCCAAGC................. 1023
     ::|||:||:: |||::||||  |||:::|
433  rgAspGlyGlyAsnArgGlyAspGlyAlaProGluArgGlySerGlyLeu 449

1022 ...GGGCCCGGCAATGGACACGAGTGGGGCCGACAGGAGGGGCCGACAGC 976
        |||||||  |||   ::|||
450  ProGlyProGlyCysGlySerProGlyLeuGlySerProGlySerProProAl 466

975  C.......................................CTTCTGCTGG 965
     |
466  aGlyTrpGlyGluProGlyGlyAlaArgGlyProGlyProGlyAlaAArgGluArgAlaA 483

964  CTCGGTGGGGCCCAGCGCTGCCT...........CCTCAGCCACCAGCA 927
     ::|||       |||
483  laArgProGlyValArgGlyValProArgProGlyGluProValProProGly 499

926  GTGTGGCTGCTACCGCAGGTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGG 877
                                 |||  :::||  |
500  ................................GlyMetGlyGlyThrGlyGl 506

876  CACTCCTCCT.............GGGTCCCCAGGTAGGGGCCAGGGGC 842
     |||  |||       ||||||||::||||||::|||||   |||
506  yThrGlyProArgSerGluGlyAlaGlyCysProAlaArgGlyAlaGly. 522

841  ACTGGTGTCCCAGTCAATGGCAGGCAGGA..................... 813
         |||   ||||||
523  ........ProProAlaTrpGlyAlaGlyProProArgAspGlyArgGlyGly 536

812  ...GGTAGCCCAGGCAGCCCCAAGACTGATCATGAAGGCATAGACAGAG 766
        :::  ||:::|   |||
537  AsnArgGlyAspGlyAlaProGluArgGlySer............... 547

765  TAGGCCTGGCGACAGTGGTCCGGGT........CCCGGAAGAGGTCAGA 725
       |||      ||||||   |||   ::|:|
548  .GlyLeuProGlyProGlyCysGlySerProGlyLeuGlySerA 562

724  GAGCAGGGCCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCA 675
     ||:::  ||||||:::  ||:  :::
562  rgSer...ProProAlaGlyTrpGlyProGlyArgGlyProGly 577

674  ......GCCCCAGGCCAGGATGAGAGCAGTGCCAGCTCCAGG.. 639
            |||||
578  AlaArgGluArgAlaAlaArgProGlyValArgGlyValProArgProGlyG 594

638  .GCCTGGGATCCGGGCACAGCAGCCCTGCTAGCCACCCGGCCCCTGGGAT 590
       ||||||
594  uProValProProGly................................ 599

589  GAGAAAGAGGCTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAAGGGCC 540
     |||       ::||||   |||||
600  ........GlyMetGlyThrGlyThrGlyThrGlyProArg........ 610

539  GGCGGCGGCCATAGCGTCCACGCGTCCAGTGGTCACTGGCCTAGGAGC 490
                                        ||| |||||
611  ............................SerGluGlyAl 614

489  GGGACACAGACCAGGCCACACTGGACCAATGCCCAGCACCATGTCAT 440
     ||||||||
614  aGly.......................CysProAla........ 618

439  GAACTTCTCCTCTACCCCACTTCCAGCAGCAGAGCGGCACATAGGTGA 390
     :::
618  ...................................SerGluGlyAl... 618

389  TGCCTGCGGCAAACACACCTCAGGCCAAAGGTTAGCAGGTTGACCAGC 340
     |||||:::        |||||||||::: |||       |||
619  ......ArgGlyAlaGlyProProAlaTrpGly...AlaGlyProProAr 632
```

```
339  AAGAGCTGGCTTCCGCTGCCGCCAGCAGGCGGCTCACCCACAGCCTCTG 290
     |||       |||    :::  |||:|||  |||:|||
632  gArgAspGly.....GlyAsnArgGlyAspGlyAlaPro........... 643

289  GACCCATAGTGGCCAGGCGGGTAGGGCTCAGGGGGCCGTTCAGGCACTCC 240
     |||             ||||||  |||||
644  .........GluArgGlySerGlyLeuProGlyPro............. 652

239  AGAACTGCTTCTGTCTCCGGCTCTGCTCCAGAAGCTGCGGCCTCTCCTCCTT 190
     :::|||||            |||||||  |||||
653  ............................GlyCysCysGly.SerPro.... 657

189  GCTGCGGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTCTGCCCAGCCC 140
     |||         |||   :: :||:|||
658  ........GlyLeuGlySerArgSerProProAlaGlyTrpGlyGluPro 671

139  TTTGGTGCCGTCCA....................GCTTCTCAGCCCATGCTCAA 105
     ||| ||||||                  ::|||:|||:|||  ::|||||
672  GlyGlyArgGlyProGlyAlaArgGluArgAlaAlaArgProGlyValAr 688

104  CACTCTGCTCGTGTGGGCACCTCAGTGGGGA 74
     :::|  |||:::  |||
688  gValProArgProGlyGluProLysArgGly 698
```

seq_name: sp_invertebrate:Q26634

seq_documentation_block:
```
ID   Q26634          PRELIMINARY;      PRT;    1414 AA.
AC   Q26634;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ALPHA-1 COLLAGEN.
GN   COLP1ALPHA.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinoidea;
OC   Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 9234811.
RA   EXPOSITO J.Y., D'ALESSIO M., SOLURSH M., RAMIREZ F.;
RT   "Sea urchin collagen evolutionarily homologous to vertebrate
RT   pro-alpha 2(I) collagen.";
RL   J. Biol. Chem. 267:15559-15562(1992).
DR   EMBL; M92040; AAA30035.1; -.
DR   PFAM; PF01410; COLFI; 2.
DR   PFAM; PF01391; Collagen; 17.
KW   Extracellular matrix.
SQ   SEQUENCE   1414 AA;  133024 MW;  5B68B99A CRC32;
```

alignment_scores:
```
        Quality:  364.00                Length:  1224
        Ratio:      0.729                Gaps:      85
Percent Similarity:   40.768    Percent Identity:  26.225
```

alignment_block:
```
US-09-030-606-110/rev x Q26634
Align seg 1/1  to:  1  from:  1 to: 1414

3029 GGGAAGTTGGGGTAGGGGAAAGTTGGGGTAGGGGAAATTTGGGCAG 2980
     |||  |||     :::|||  :::|||  |||  ||:||
43   GlyAspLysGlyGlnLysGlyGluProGlyAspAlaAspIle...AsnSe 58

2979 TGCCTTCATCAGCCCAGTCCTAGCAGAGAGAGGGG.......... 2943
     |||   |||||
58   rAlaAsnPheProProGlyLeuProGlyProValGlyProProGlyProS 75

2942 ..........AGTGGAAGTGGGGGGAACCAGGCT........GGG 2916
     ||||  |||||::||||||  |||
75   erGlyProSerGlyProAlaGlyAsnAsnGlyProProGlyProAsnGly 91
```

```
2915 CCAAGA..........GAAGAGGGGTGGTTAGGAAGCCGTTGAGACC 2878
     ||||||          :::||||| |||          || |||
92   ProArgGlyAsnProGlyMetAspGlyLeuThrGlyLeuProGlyIlePr 108

2877 TGAAGCCACCCTCTACCT..................... 2858
     :::||||||:::: |||
108  o...GlyProProGly.ProProGlyLysSerGlySerLeuValAlaSer 123

2857 ...........TCCTTCAACACCCTAACCTTGGGTAACAGCATTTGGAA 2820
                |||||||||    |||:::         |||||
124  AlaGlnThrSerSerPheAsnLys.........GlyProSerLeuAlaGl 137

2819 TTATCATTTGGGATGAGTAGAGAATTTCCAAGGTCCTGGGTTAGGCATTTG 2770
     :|||    ::::|||:::::||        ::  |
137  yTyr..........GlnTyrProGlnAla......GlnAlaAlaG 147

2769 GGGGGCCAGACCCCAGGAGGAGAAGAGATTCTGGCAATGATCAGCCCAATGA 2720
     ||| |||  ||||||
147  lyThrProGlyProArg................................. 152

2719 CCAGCTATCTCAGGGGACCTGATTGTTGGGGATCCCCCACCCTACCCAAA 2670
     |||||||   |||||     |||||| ||||||
153  ...................GlyProGlyProProGl 159

2669 TATTAGACACCAACAGAAAAGCTAGCAATGGAATCCCTTCTACTTTGT 2620
     |||  :::  |||||    ::: |||||||
159  ySerArgGlyProGlnGlyLeuThrGlyProSerGlyProSer 173

2619 TAAATAAATAAGTTAAATATTTAAATGCCTGTGTCTCTGTGATGGCAACA 2570
     :::||||||||    ::||
174  .................GlyGluThrGlyProSerGlyAsn 181

2569 GAAGGACCAACAGCCACATCCTAGAAAAGGTAAGAGGGGGTGGATCA 2520
     :::||||||| |||      |||:::    ||| |||
182  SerGlyProProGlyProSerGlyLeuProGlyArgProGlySerAspGl 198

2519 GCAAAAAGACAGTGCTGT.......GGGCTGAGGGGACCTGGTTCTTGTG 2477
     ||||||||  :::       |||  ||||||||||:::::|
198  yAspAspGlyThrProGlySerGlnGlyValGlnArgGlyProAlaGlyThrP 215

2476 TGTTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAATCC 2427
     |||||||   |||                    ::::|||
215  roGlySerArgGlyThrProGlyMetProGlyIleAlaProGlyMetLysGly 231

2426 CATGGAGGAGTGTTTCATCCTAGAAAACTCCCATGCAAGAGCTACATTAAA 2377
     ::::|||||| |||    :::::||  ::::|||
232  HisGlnGlyLeu......ProGlyMetThrGlySerLysGlyGluArgGl 246

2376 CGAAGCTGCAGGTTAAGGGGCTTAGACGATGGGAAACCAGGTGACTGAGTT 2327
     :::||| ||||||    |||||||||::::::
246  yGluGlyGly.GluArgGly...SerAspGlySerProGlyProValGly 261

2326 TATTCAGCTCCCAAAAACCCTTCTCTAGGTGTGTCTCAACTAGGA...... 2282
     ::::|| ||||||||:::||||| ::::|||
262  AlaProGlyProAlaGlyLysProGlyGlnProGlyGluArgGlyArgTh 278

2281 ..............................GGCTAGCTGTTAA 2269
                                   |||    |||
278  rGlyProAlaGlySerGlnGlyAspArgGlyAlaAspGly.....AlaT 293

2268 CCCTGGCCTGGTAATCCA...............CCTGCAGAGTCCCCG.... 2234
     ||| |||||          :::       |||::::
293  hrGlySerGlnGlyProProGlySerThrGlyProAlaGlyProGly 309

2233 ...................CATTCCAGTGCATGGAGCCCTCT..... 2210
                        ::::|||     |||||||
310  MetProGlyIleSerGlyAlaLysGlyAspAlaProGlySerProGlyAlaAr 326

2209 .GGCCTCCCTGTATAAGTCCAGACTGAAACCCCCTTGGAAGGCCTCCAGT 2161
     |||||||  ||||:::      |||||||:|
326  gGlySerPro...........GlyLeuGlnG 333
```

```
2160 CAGGCAGCCCTAGAGACTGGGAGAGG..........AGAGGG 2126
     ::::| |||||||                      |||||||
333  lyAlaArg..........GlyGluArgGlyGluGlySerGlnGly 345

2125 ACGGCCCAGCCCCAGCTGTGCAGCTACGC............ 2096
     |||||||:::|| |||
346  GlnThrGlyProGlyValProGlyTyrArgAspGlySerAsnGlyAlaLy 362

2095 ...........ACCTCAGCAGCAGGGT.........GGCA 2074
     :::||||||||   |||:::     |||:
362  sGlySerAlaGlyProSerGlyAlaGlnGlyThrProGlyPheProGlyA 379

2073 GCAGA..................GAGCCACATTACTTGGCAGCAAC 2045
     ::||                    ||||||  |||||||
379  laArgGlyProProGlyProAlaGlySerProGlyProAlaGlySerLys 395

2044 AGAAACTGGCGGCCAGCCCGCAGCCCCATGGGCTAACAGGAGCGGGGA 1995
     |||                   :::    ||||||||||||||
396  GlyAspGlnGlyAsnProGlyGlnProGlyAlaGlnGlyGluArgGlyPr 412

1994 GCTGGGACCC...AGTGAGGCAGGCCCT.....CCACCCCAATGTGCTG 1954
     ||||||||||   :::||::::|||||      |||||
412  oLeuGlyProArgGlyGluThrGlyProAlaGlyProPro........ 425

1953 GAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTCAAATACTAC 1904
425  .................................................. 425

1903 CTGTGTAGCAAAGTAAATGGCACCAGGACCCCAGGCCTGCGGCAGCACCA 1854
     :::   :||||||||    ::  :||
426  ........GlyAlaGlnGlyGluSerGlyGluArgGlySerArgGlyAl 439

1853 TATAGGCAGTGA...CAGACTGGCTGAGCTGGACAATGAGCCCCATAAA 1808
     ::||||        |||    |||||||||  :::|||
439  aLeuGlyProAlaGlyProProGlyValGlyGluArgGlyPro..... 454

1807 CAGGGATGGGGCCACCTGGGACAGCAGGA.............. 1779
     ||||||                :::|||
455  ....MetGlyProProGlyMetSerGlyAlaProGlyAlaProGlyAla 469

1778 ..........AGGCACTATCCAGGATGGCGAGT.......... 1755
              |||            :::|||||
470  LysGlyAspArgGlyLeuProGlyGluArgGlySerAlaGlySerLysGl 486

1754 ...........CCAGGCAGATGCCCCGGCCCGGAACCACCTGCCCTC 1718
              |||         |||  :::||||||||
486  ySerAlaGlyLeuSerGlyArgProGlyLyProGlyMetProGlyGln 502

1717 GGTGGGCCTCACCCACCACCACGTACGGAGACATCACAGGCAGGAGGCCC 1668
     |||||||||                         ||||
503  ArgGlyLeuThr.................................GlyPr 508

1667 CGCAGAGCGCGGTGGGGTGAGGTGGGAGCAGGCCACTGCCT.......... 1630
     | ::::  |||                 :::|||| |||
508  oProGlyLys.GlnGlyTyrArgAspGlyLysProGlyProAlaGlyAlaPro 524

1629 ......CCAGCACCCACGTGCCATTAGGG...........AAGGGAGC 1598
          |||||      ::: |||  |||             :::|||
525  GlyGluProGlyAsnSerGlyProAlaGlyAlaSerGlyGlnArgGlyLe 541

1597 TCCAGGCTTA..............GGGCCT.................. 1582
     |||:::                  ||||||
541  uProGlyLeuValGlyLeuProGlyProGlnGlyArgGlyGluArgG 558

1581 .........GCGAGGAAGCTGGTCATCAGCCTGTCCTCACTGCTAGCACCT 1540
              ||| ::::     |||||||||||||
558  lyGluAspGly.GlyGlnGlySerProGlyAlaPro............ 569

1539 CCAGTGTCCCCTCGGTATTTGGGCAGGAACACCTGCTTCTCCCGGTGGTA 1490
     :::::||||||| ||||||                |||||||
570  ........GlyLeuThrGlyGluProGlyLyLysA 578

1489 GAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGGCTGAGAAGGTGAACC 1440
```

Wed Sep 29 14:26:57 1999

```
  788  roGlyProGlnGlyGluLeuGlyProThrGlyAlaArgGlyGluThrGly  804
       |||||:||              ||  ||||||
  665  ...CCAGGATGAGCAGTGCCAGCTCCAGGGCCTGGGATCCGGCACAGC  619
                              :::   ||||||||
  805  GlyProGly........ProSerGlyProThrGlyAspProGlyProGl  818

  618  AGCCCTGCTAGCCAGCCCGGCCCTTGGATGAGAAAGAGGCTCAGGCAGGAT  569
       :::  ||||| :::|:::   ||||||||  ::||
  818  nGlyProLeuGlyAlaProGlyGlnGlnGlyLeuArgGlyGluThrGly.  834

  568  GCCCAAGGACAGTGCCCAGATGAAGGGCCGGCGGGCCATAGCGTCCAC  519
       ||:::|
  835  ........ProGln  836

  518  GCCAGTGGTCACTGGCTGAGCCTAGGAGCGGGACACAGACCAGGCCCAGC  469
                          ||  |||   |||||||
  837  ........GlyGlnGlyGlyProProGlyPro.  844

  468  ACTGGACCAATGCCACCATGGTCATGAACTTCCCTCTACCCCCAC  419
       ::    ||||||   ||||||
  845  .IleGlySerLeuGlyAlaProGlyAlaGlnGlyLeuProProGlyProThrG  861

  418  TTCCAGCAGCAGGGCCGCACATAGGTGATGCCTGCGCCAAACACACCT  369
       ||:::::|||::|||||         |||:::  |||
  861  lyProSerGlyAsnAlaGly.......SerProGlyGlnPro  872

  368  ........CCAGGCCAAAGTTAGCAGCAGCAAGAGC  334
                  ||||||||
  873  GlyAlaArgGlyProGlyGln......SerGlySerProGlyGlnPr  887

  333  TGGGCTTCCGGTGCCCAGCA.......GGCGGCT  305
       |||||:||||         ||
  887  oGlyLeuAlaGlyArgThrGlyProSerGlyGluArgGlyAspLysGlyA  904

  304  CACCCACAGCCTCTGGACCATAGTGGG...CCAGGCGGTAGGCGTCAGG  258
       ::              ||||||||   ||  ||||||  ::
  904  snAspGlyGlnSerGlyPro.ProGlyProProGlyProAlaGlyProAl  920

  257  GGGCCGTTCAGGCACTCCAGAACTGCTTCGTCTC......GGCTCTGCTC  214
       :||:||||||  |||          :::||  ||:::::|
  920  aGlyGlnSerGly.......IleLeuGlyLeuAlaGlyLySerGlyP  934

  213  CAGAAGCTGCGGCCTCCT.................  194
       ||::::::  :|:::|||
  934  roArgGlyProGlyLyProAlaGlyProAlaAlaGlySerArg  950

  193  ...CCTTGCTGCCCAACTGCCTAGGAATCAGCCAGGCGCCCATTTCTG  147
       ||:::|                  ||:|||||||||
  951  GlyProAlaGlyLysSerGlyLyAspArgGlySerProGlyAla........  964

  146  CCAGCCCTTGTGCCGGTCCCAGCTTCTCAGCCCATGCTCAACACCTGCT  97
       :||||||||||:::|||:::||              :||  |||||
  965  oGlnGlySerArgGlyLyLysGlyAspThrGlyAlaSerGlyAlaAsnG  974

  96   GCTGTGGGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCTGGC..  49
       :::  ||||              ||  :||:::||||||||||
  975  GlyGluAsnGlyMetProGlySerAspGlyAsnAspGlyAlaProGlyPr  991

  48   ...CCAGGCGCGGCCGGCTGTCACCCGGAGCCAGCCGCCTGCGAGGC  9
                ||||||
  991  oGlnGlySerArgGlyLyLysGlyAspThrGlyAlaSerGlyAlaAsnG  1008

  8    TGGTTCCC  1
       |||
  1008 lySerPro  1010
```

```
  578  !||||:::||          |||  ||||||
       rgGlyGluPro..........GlyValAlaGly.  585

  1439 CGGTGAGGGGCCTGAAGCTGTCACCACGGCCACACTGTGGACAGGCAT  1390
       |||  |||:::|||  ||:::||||
  586  .........ProProGlyProGlnGlySerAlaGlyI  595

  1389 GTGGCACCGGCAG...CCACAGGGAAAGTCTGCCCACACTGGCCAAATAGAC  1343
       ::|||  |||||||||
  595  uArgGlyAsnGlnGlyProGlnGly.........  603

  1342 TGCTCGAGTGCCGAATCGGTGCCACCAGCCGGTCATGACCAGAGAGAAGA  1293
       ::|||  :::|||
  604  ....GlnAlaGlySerMetGlyProProGlyPro.....  613

  1292 CCAGGGAGATGGCGCACTGCAGGAACAGCC......CCAGGCTGCCCATC  1249
       |||||                  ||  |||  |||:
  614  ProGlyAlaSerGlyAspAlaGlnGlyAlaGlyGlnGlyAspAsnGlyProProGlI  630

  1248 CGAACGCCTTCATCATATGTCTCCGGCCCTCGGTCCCGGCTCAGCTCT  1199
       :|||              |||||||||:::  |||
  630  yGlu..........SerGlyProGluGlyPro..........  637

  1198 GGGCACGCCCTGGTACA.........GCCCCTCGCCCCACGAAAT  1164
       |||:::           |||||  |||:::   |||
  638  ........ProGlyAlaAargGlyGluArgGlyAlaProGlyGluArgGly  651

  1163 CCGTGTAAAAACAGCGTGAAGCTCATGATGCCATCCAGCTGCCACAGCTCA  1114
       ||              :::  |||  |||||||
  652  Pro...............GlnGlyLeuThrGlyAlaGlI  659

  1113 GCCACGAAGAGCC...GGCCAGGGTGCGGGGCATGCCGGCCAGCACAGCT.  1068
       ||:  ||||:::  :::||||  |||:::|||||||
  659  nGlyArgArgGlyIleGlyGlyArgAlaGlyAlaGlyIleAsnTyrGlyAlaThrGlyP  676

  1067 ........GGTGCAGCCGGGGAGCAGGGGCGCCCAGGTTCC..........  1035
       |||  ||||||  |||||:::
  676  roProGlyGlnLysGlyGluMetGlyProProGlyProGlyAsnValGlyLeuGln  692

  1034 ..........GGAAAG....CCAAGCGGGCCGGCCATGGACAGCAGTGG  1000
       |||:::     |||:::
  693  GlyProProGlyGluLeuGlyProSerGlyPro............ProGlI  705

  999  GGGCGACAAGGAGGGGGGCCGACGCCCTTCTGCTGGCTCGGGTGGGGCCCAG  950
       |||              ||||||||
  705  yAla.......ArgGlyProPro.............GlyPros  713

  949  CGCTGCCTCCTCAGCCACCAGCAGTGTGGCTGCTACCGCAGGTGAGGAAGA  900
       ||:  |||:::
  713  erGlySerProGlyProAsp.............GlyPro  721

  899  TGAGGGTGAGCCAGGCAAAGAGGCACT.........  873
       ||||  ||||:::||||||
  722  AlaGlyAlaGluGlyAspArgGlyProValGlyProMetGlyProSerGI  738

  872  .CCTCTGGGTGCCCAGGT..AGGGGG...........CCAGGGCACT  839
       |||  |||  ||||||              ||||||
  738  yProSerGlyMetProGlyValArgGlyGluArgGlyAspAsnGlyGluProGly....  753

  838  GGTGTCCCAGTCAATGGCAGGAGGTAGCCCAGGCAGCCCCCCAAGAC  789
       :::||||||||||                  |||
  754  ..........ProGlyGlyAlaProGlyGlnArgGlyAsp  763

  788  TGATCATGAAGGCATAGAACAGCAGTAGGCCTGGCAGCAGTGGTCCG....  744
       ||||||||  |||||   ||||||
  764  ................ProGlyAspLeuGlyProGlnGlI  771

  743  .GGTCCCGGAGAGGTCAGAGAGCAGGCAGGGCCTTCCAGTGGAGTGAAGCACAC  695
       ::::||||||      |||||:::   ::::::|
  771  ySerProGlySerProGlyPheAlaGlyLyProProGlyArgSerGlyAsnP  788

  694  CTGGCCACAGAAGTCCAGCAGCCCCACGC.  666
       |||||  :::::||||
```

```
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 69.4 KD PROTEIN.
GN   GAMMAHV.M10C.
OS   murine herpesvirus 68.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WUMS;
RX   MEDLINE; 97366649.
RA   VIRGIN H.W. IV., LATREILLE P., WAMSLEY P., HALLSWORTH K., WECK K.E.,
RA   DAL CANTO A.J., SPECK S.H.;
RT   "Complete sequence and genomic analysis of murine gammaherpesvirus
RT   68.";
RL   J. Virol. 71:5894-5904(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WUMS;
RA   LATREILLE P., WAMSLEY P., WATERSTON R.H.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97553; AAB66422.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   727 AA;  69448 MW;  2391981E CRC32;

alignment_scores:
         Quality:   359.00              Length:     816
         Ratio:       1.072             Gaps:        54
Percent Similarity:   41.054     Percent Identity:   28.064

alignment_block:
US-09-030-606-110/rev x 041973     ..

Align seg 1/1  to: 041973  from: 1  to: 727
```

```
2181 CCCCTTGGAAGGCCTCAGTCAGGCAGCCTAGAGACTGGGGAGAGAGG 2132        129 yProProArgArgAspGly.GlyAsnArgGlyAspGlyAlaProGluArg 145
     |||||| ||:::  :::|||::::|| :::                                 |||||| ||     ::     ::   |||||||||:::::
  30 ProProArgArgAspGlyGlyGlyAsnArgGlyAspGlyAlaProGluArgGl 46   1747 GATGCC........CCGGCCCGGAACCACCTGGCCTCGGTGGGCT 1710

2131 AGAGGGACGCCCCAGCTGTGCAGCTACGACCTCAGCAGCACA 2082          146 GlySerGlyLeuProGlyProGlyCysGlySerProGlyLeuGlySerAr 162
     |:::||   |||::||:::|||:::||::: |||::|||                        ::        :::    |||||||||||:::
  46 ySerGly.LeuProGlyProGlyCysGlySerProGlyLeuGlySer... 61  1709 CACCCACCACCACGTACGGAGACATCACAGGCAGGAGGCCCCCAGAGC 1660

2081 GGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAACTGGCGGC 2032   162 gSerProProAlaGlyTrpGlyValArgGlyGlyProGlyAlaA 179
     |||||| |||||| ||  :::|:| :::|||||                              :::                  |||||::||:::::    |
  62 ..........................Ar 62              1659 GC................GGGTGGAGTGGGAGCAGGCCACTGCCTCC 1628

2031 CAGCCCGCAGCCCCATGG...............GGCTAAC 2006            179 rgGluArgAlaAlaArgProGlyValArg.ValProArgProGlyGluPr 195
     |||||  ||| |||  ||                ||:::                        ||         |||::||  |||   ||||
  62 gSerProProAlaGlyTrpGlyProGlyGluProGlyGlyAlaA 79   1627 AGCACCCACGTGTCCATTAGGGAAGGGAGCTCCAGGCTTAGGCCTGGCA 1578

2005 AGGAGCCAGTGAGGCAGGCCCCTCCACCCCA 1962                     195 oValProProGlyGlyMetGly..GlyThrGlyGlyThrGlyPro...A 210
     ::||||||::::||||||||||::: |||||                               |::||| ||:::||:|||||||||  ||  |||||
  79 rgGluArgAlaAlaArgProGlyValArgGlyValArgProGlyGluPro 95   1577 GGAAGCTGGTCATCAGGCTGTCCTCACTGCCTAGCACCTCCAGTGTCCCT 1528

1961 ATGTGCTTGGAAGTTTCTTACGCTGAGATATTGGCCAAGTCGCTCTTGTCA 1912   210 rg.SerGluGlyAlaGlyCysProAla.......ArgGlyAlaGlyProPr 224
     ::|                          :::                |||| ||||:::: |||||      :::
  96 val............................. 96         1527 CGGTATTGGCAGGAACACCTGCTTCTCCCGGTGGT...AGAGGGAGGC 1481

1911 AATACTACCTGTGTAGCAAAGTAAATGGCGACCAGACCCAGCCTGCGGC 1862    224 oAlaTrpGlyAlaGlyLyProProArgArgAspGlyGlyAsnArgGlyAspG 241
     |||||||||||| |||  ::|||||::|||||                              |||||| ||||| |||   |||||   |||||
  97 ........ProProGlyGlyMetGlyG 103             1480 CAGTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGG 1431

1861 AGACA..........................CCATATAGG 1848           241 lyAlaProGluArgGlySer..GlyLeu.......... 249
     |||                            |||  :::                        ::: |||||  |||||
 103 lyThrGlyThrGlyProGlySerGluGlyAlaGlyCysProAla.ArgG 119  1430 CGGGTGAAGCTGTCACCACGGCCACACTGGGACA...GGCATGTGGCA 1384

1847 CAGTGACGACTGGCTGAGCTGGACAATGGAGCCCATAAACAGGGATCGG 1798   250 ............................ProGlyProGlyCysGlySerProGlyLeuGlySe 261
     ||| ||| :::||::|| ||                                          CCGGCAGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT 1334
 120 .........GlyAlaGlyProProAlaTrpGlyAlaGl 129  1383 CCGGCAGCCACAGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAGT 1334
                                                              |||:::||||
                                                         261 rArgSerPro.................... 264

1797 GCCACTGGGACAGCAGGAAGGCCACTATCCAGGATGGCGAGGTCCAGGCA 1748   264 ..................... 264
     ||:::|||      |||||                                     1333 GCCGAATCGCTGCCCAGCCGGTCCATGACGAGAGACCAGGGAGA 1284
 361 rArgSerProProAlaGlyTrpGluProGlyArgGlyProGlyA 378          |||||||||
                                                         264 ...................ProAlaGly.................... 267
```

```
1709 CACCCAC ... (continued columns)

1283 TGG...CCGACTGCAGGAACAGCCCAGGCTGCCATCCGAACGCCTTCA 1237
     ||| ::: :::||||:::||||||:::                           282
 267 ................. 267                         268 TrpGlyGluProGlyLyArgGlyProGlyAla...ArgGluArgAla. 282

1236 TCATAGTGTCTCCGGGCCTCCGGTGCCCGGTCAGCTCTGGCACGCCCTG 1187
     ||||||||                                              286
 282 ................AlaArgProG 286

1186 GTACA...GCCCCTCGCCCACGAAATCCGTGTAAAACACGCTGAAGGTCA 1140
     ::||| ::||||||||| :::|||                              295
 286 lyValArgValProArgProGlyGluPro........... 295

1139 TGAGTGCCA..........................TCCAGCTGCACAGCTCAGCCACGAAG 1105
     |||||                          ::: ||||||                311
 295 ....ValProProGlyGlyMetGlyGlyThrGlyThrGlyGlyThrGlyProArgSe 311

1104 AGCCGGCGCAGGGTGCGGGCA...........TGCGGCAGCACAGCTGGTGCAG 1061
     ||||||||||| |||       :::||                             328
 311 rGluGlyAlaGlyCysProAlaArgGlyAlaGlyProProAlaTrpGlyA 328

1060 CCGGG...GAAGCAGGGCGGCCCAGGTTCCGGAAAGCCAAGC 1023
     :::|| ::||||||:::||:||| |||:::||||                      344
 328 laGlyProProArgArgAspGlyGlyAsnArgGlyAspGlyAlaProGlu 344

1022 ..........GGGCCCGGCATGGACGAGCAGTGGGCGACAAG 991
     ||||||| ||||| ||                                       361
 345 ArgGlySerGlyLeuProGlyProGlyCysGlySerProGlyLeuGlySe 361

990 GAGGGGCCGACAGCC.......................... 975
    |||:::|||  |||                                          378
361 rArgSerProProAlaGlyTrpGluProGlyArgGlyProGlyA 378
```

178 CTGCCTAGGAATCAGCCAGCGCCCATTCTGCCAGCGCCTTGGTGCC... 131
          ||| ::: ||||||
629 GlyProProArgArgSerGlyGlyAsn......ArgGlyAspGlyAlaPr 643

130 ................................GGTCCAGCTTCTCAGCCCATGCTCAACA 103
                                      ||||||:::
643 oGluArgGlySerGlyLeuProGlyProGly.............. 653

102 CCTGCTGCTGTGGGGCACCTCAGTGGGACACGTCTCA 65
        |||||||||| || |||
654 ....CysGlySerProGlyLeuGlySerGlnSer 663

seq_name: sp_rodent:Q63079

seq_documentation_block:
ID   Q63079   PRELIMINARY;   PRT;   1453 AA.
AC   Q63079;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   COLLAGEN ALPHA1 (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE OF 1-1092 FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=TOOTH;
RA   BRANDSTEN C., LUNDMARK C., CHRISTERSSON C., HAMMARSTROM L., WURTZ T.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z78279; CAB01633.1; -.
DR   PFAM; PF01410; COLFI; 1.
DR   PFAM; PF01391; Collagen; 18.
DR   PFAM; PF00093; vwc; 1.
DR   PROSITE; PS01208; VWFC; 1.
FT   NON_TER   1.
SQ   SEQUENCE   1453 AA; 137886 MW; 845910AE CRC32;

alignment_scores:
                Quality: 358.00   Length: 1045
                  Ratio: 0.810   Gaps: 67
Percent Similarity: 42.297   Percent Identity: 27.081

alignment_block:
US-09-030-606-110/rev x Q63079

Align seg 1/1 to: Q63079 from: 1 to: 1453

2594 TGCCTGTGT...................CCTGTGATGGC.........AACAG 2569
         |||:||| ::|||||||| :::||
52 CysIleCysHisAsnGlyThrAlaValCysAspGlyValLeuCysLysGl 68

2568 AAGGACCAACAGGCCACCATCCTGATAAAAGGTAAGAGGGGGTGGGATCAG 2519
       : |||::||||:|||::||||||:||||||| :|||:|
68 uAspLeuAspCysProAsnProGlnLysArg...GluGly.......... 80

2518 CAAAAAGACAGTGCTGT..................................... 2502
       ::|||||
81 ....GluCysCysProPheCysProGluGluTyrValSerProAsp 94

2501 ............................GGGCTGAGGGGACCTGGTTCTTGTGTTGCCCCTCAG 2464
                                      ::||||||||
95 AlaGluValIleGlyValGluGlyPro........................ 103

2463 GACTCTTCCCCTACAAATAAGTCATATGTTCAAATCCCATGGAGGAGTGT 2414
                     |||:|||||||:|:|||||||:||||| :||||
104 ....LysGlyAspProGlyProG 110

2413 TTCATCCTAGAAACTCCATGCAAGAGCTACATTAAACGAAGCTGCAGGT 2364
       |||||||::|:: |||::
110 lnGlyProArgGly.ProValGlyPro..................... 118

2363 TAAGGGGCTTAGAGATGGAAACCAGGTGACTGAGTTTATTCAGCTCCCA 2314

```
 376 AsnProGlyAlaAspGlyGlnProGlyAlaLys.GlyAlaAsnGlyAlaP 392
1555 CTCACTGCTAGCACCTCCAGTGTCCCCTCGGTATTGGCAGGAACACCT 1506
                   ||||||     |||||::|||
 392 ro...................................GlyIle..AlaGlyAlaPro 398
1505 GCTTCTCCCGGTGGTAGAGGAGGCCAGTGTGTAGGCAGGATCTGCAGG 1456
         :::||| ||||||:||:|||
 399 GlyPheProGlyAlaArgGly...........ProSerGlyPro.GlnG 411
1455 GCTGAGAAGTGAACCCGGTGAGGCGGCTGAGAGTCGTCACCACGGCCAC 1406
     ||  ::||| |||||||  |||   :::|:|||
 411 lyProSerGlyAlaProGlyPProLysGlyThrSerGlyEluProGly... 426
1405 ACTGTGGGACAGGCATGTGCCAGCCACCGGCAGGGAAAGCTGCCACAC 1356
     ::: |||:| |||| |||||:::  ::||:||
 427 AlaProGlyAsnLysGlyAspThrGlyAla.LysGlyGlu......... 439
1355 TGGCCAAATAGACTGCTCGAGTGCCAGATCGTCGACCAGCCGGT..... 1311
                          |||||:|||
 440 ....................................ProGlyProAlaGlyValGl 446
1310 .......CCATGACCAGAGAGAGACCAGGAGATGGCGCACTGCAGGAA 1268
                ||| |||              ::|
 446 nGlyProProGlyProAlaGlyEluGluGlyLysArgGlyAlaArgGlyGlyG 463
1267 CA.....................GCCCCAGGCTGCCCATCCGAACGC....... 1242
       |||||||||||||:||||||
 463 luProGlyProSerGlyLeuProGlyProProGlyEluArgGlyGlyGlyPro 479
1241 ........CTTCATCATAGTGTCTCCGGGCCTCGGTGCCC 1210
 480 GlySerArgGlyPheProGlyProGlyAlaAspGlyValAlaGlyProLysGlyProPr 496
1209 GGCTCAGCTCTGGCAGCCCCTGTACAGCCCCTCGCCCACGAAATCCGT 1160
     ||| |||||  :::|||    :::||
 496 oSerGlyEluArgGlySerProGlyProAlaGlyProLysGlySerPro. 512
1159 GTAAAACAGGCTGGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCAGCCA 1110
 512 ................................................. 512
1109 CGAAGAGCCGGGCCGGGGCCCAGGCGTGCGGGGCCATGCGGCCAGCCACCAGCTGGTGCAGC 1060
                                                ::: ||||||
 513 ...................................GlyGluAlaGlyArgPr 518
1059 CGGGGAAGCAGGGCGCCCAGGT.......TCCGAAAGCCAAGGC 1022
     |||||||  ||||||       ::|:|:::
 518 oGlyGluAlaGlyLyLeuProGlyAlaAlaLysGlyLeuThrGlySerProGlyLys 535
1021 GGCCCCGGCCATGGACAGCAGTGGGGCACAAGGAGGGGGCCGACAGCCCTT 972
     ||||||| ||||| |||||| |:|::||:|||||:|
 535 erProGlyProAsp......GlyLysThrGlyProProGlyProAlaGly 549
 971 CTGCTGGCTCGGTGGGGGCCCAGCGCTGCCT........... 942
     ::||| |||||||||| |||
 550 GlnAspGlyArgProGlyProAlaGlyProProGlyAlaArgGlnAl 566
 941 .................CCTCAGCACCAGCAGTGTGGCTGCTACGCAGGT 908
                      ||| |||              :::::
 566 aGlyValMetGlyPheProGlyProLysGlyGlyThrAla.......... 578
 907 GAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCCACTCCT........... 870
     |||||||||||||||||  ||||
 579 .GlyGluProGlyLysGlyAlaGlyArgGlyValProGlyProProGly 594
 869 ....................CCTGGGTGCCC 859
 595 AlaValGlyProAlaGlyLysAspGlyGluAlaGlyAlaGlyGlyAlaPr 611
 858 AGGT...AGGGGCCAGGCACTGGTGTCCCAGTCAATGGCAGGCAGGA. 813
     ||||||:::          ::||| |
 611 oGlyProAlaGlyProGlyEluArgGlyGlyGlnGlyProAlaGlyLys 628
```

```
 119 .ProGly...GlnAspGlyIleGlyProGlyGlnProGlyLeuProGlyProP 134
2313 AAAACCCT.........TCTCTAGGTGTGTCCAACTAGGAGGCTAGCTGTTA 2270
     :::       ||||||   |||||||||
 134 roGlyProProGlyProProGlyProLeuGlyLeuGlyAsnPheAla 150
2269 ACC.....................................CTGAGCCTGGG 2256
     :::                                      ::||||||::
 151 SerGlnMetSerTyrGlyTyrAspGlYLysSerAlaGlyValSerValPr 167
2255 TAATCCACCTGCAGAGTCCCGCATTCCAGTGCATGAGCCCTTCTGGCC 2206
 167 oGlyProMetGlyProSer........GlyProArgGlyL 178
2205 TCCCTGTATAAGTCCAGACTGAAACCCCTTGGAAGGCCTCCAGTCAGGC 2156
        :::|       ||:|||           |||
 178 euPro...........................Gly 180
2155 AGCCCTAGAGACTGGGGAGGAGGAGGACGCCCCAGCCCCCAGCTGT 2106
     |||||
 181 ProPro.............GlyAlaProGlyPro........ 187
2105 GCAGCTACGCACCTCAGCACCAGAGGTGGCAGCAGAGAGCCACATTACT 2056
      :::::|||||      |||||
 188 .........GlnGlyPheGlnGlyProProGlyEluProGlyGlyEluP 200
2055 TTGGCAGCAACAGAAACTGGGCGGCCAGCCCAGCCCCCATGGGGCTAAC 2006
 200 roGlyGlySerGlyProMetGlyProProGlyProProGlyProProGly 216
2005 AGGAGCGGGGAGCTGGGA.....................CCCAGTGAGGGC 1977
     :::|:::||||  |||                       |||:::|||
 217 LysGlyAsnGlyAspAspGlyAlaAlaGlyLysLysProGlyLysArgProGlyGluAr 233
1976 AGGCCCCTCCA.........................CCCCAATGTGCTGGAA 1951
     :::|:::|||                        |||
 233 gGlyProProGlyProGlnGlyAlaAlaArgGlyLeuProGlyThrAlaGly. 249
1950 GTTTTCTACGCTGAGATTTGGCCAAGTCGCTCTTGTGCAAATACTACCTG 1901
 250 ................LeuProGlyMetLysGlyHisArgGlyPheSerGlyGly 261
1900 TGTAGCAAAGTAAATGGCGACCCAGACCCAGGCCGTGCGGCAGACACCATAT 1851
                     ::: |:|       ::|
 262 LeuAspGlyAlaLysGlyAspThrGlyAlaGlyProLysGlyGluPr 278
1850 AGGGCAGTGA.....CAGACTGGCTGAGCTGGACAATGAGCCCATAAACAG 1805
     ||||||          :::||  |||
 278 oGlySerProGlyGluAsnGlyThrProGlyGlnMetGlyPro...ArgG 294
1804 GGATG...............GGGCCACCTGGACAGCAGGAAGG 1776
     ||:::               ||||||||||||||||||
 294 lyLeuProGlyGluArgGlyLyArgProGlyProProGlyThrAlaGlyAla 310
1775 CACTATCCAGGATGGCGGAGGTCCAGGCAGATGCCCCGGCCCGAACCACC 1726
     ::: :::|||:|||  |||   ||  |||||| ||| ||
 311 ArgGly.AsnAspGlyAlaValAlaGlyAlaAlaGlyPro.ProGlyProThr 326
1725 CTGGCCTGGTGGCTCACCACCACACGTACGGAGACATCACAGGC 1676
     |||||||  ||| |||
 327 ..GlyProThrGlyProProGlyProGlyPheProGlyAlaAlaGlyAlaLysGi 342
1675 AGAGGGC...CCGCAGAGCGCGGGGTGGAGGT.......GGGAGCA 1641
     ::||||||  |||||||||:::|||            |||||||
 342 yGluAlaGlyProGlnGlyValAlaArgGlyGlySerGluGlyProGlnGlyValA 359
1640 GG.......CCACTGCCTCCAGCACCCACGTGTCCATTAGGG...AAGGGA 1600
     ||:::    |||:::||:::|||                   |||
 359 rgGlyGluProGlyProProGlyProAlaGlyAlaAlaGlyProAlaGly 375
1599 GCTCCAGGCTTAGGG......CCTGGCAGGAAGCTGGTCATCAGGCTGTC 1556
     |||||| |||                 :::::|||||:::|
```

```
  61 CTCAGATCCTGGCCGA..........................................GGCG  42
     :::  |||||||:::                                              |||
 905 aGlyArgProGlyGluValGlyProAlaGlyProProGlyProProGlyProAlaGlyG  922

  41 CGCGGCGTGTCACCCGGAGCC  22
     :::  ||||||||||||
 922 luLysGlySerProGlyAla  928

ALPHA 1 TYPE I COLLAGEN.

seq_name: sp_vertebrate:O93251

seq_documentation_block:
ID   O93251;         PRELIMINARY;         PRT;    1445 AA.
AC   O93251;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ALPHA 1 TYPE I COLLAGEN.
OS   Rana catesbeiana (Bull frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Raninae; Rana.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ASAHINA K., UTOH R., OBARA M., YOSHIZATO K.;
RT   "Spatiotemporal expression of bullfrog a1(I) and a2(I) collagen genes
RT   in intestine during metamorphosis.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB015440; BAA29028.1; -.
DR   PFAM; PF01410; COLF1; 1.
DR   PFAM; PF01391; Collagen; 18.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   1445 AA;  137252 MW;  98A415EF CRC32;

alignment_scores:
               Quality:  354.50              Length:      966
               Ratio:      0.914              Gaps:         65
      Percent Similarity:  40.166      Percent Identity:  26.501

alignment_block:
US-09-030-606-110/rev x O93251

Align seg 1/1 to: O93251   from: 1 to: 1445

2588 TGTCTCTGTGATGGCAACAGAGGAGGACCACAGGCCACATCCTGATAAAGG 2539
     |||||||||||||||||||||||||||||||||||||||||||||||||||
  33 CysValValAspGly.........................ArgThrTyrAsnAspLys.. 43

2538 GTAAGAGGGGGGTGGATCAGCAGCAAAAAGACAGTGCTGTGGCTGAGGGGAC 2489
         |||    |||    :::::|||
  44 ........AspValTrpLysProGluAlaCysGlnIle......... 53

2488 CTGGTTGTTGTGTGCCCCTCAGGACTCTTCCCTACAAATAAGTCAT 2439
     ||||||||| |||:::||| ::::|||
  54 ........CysVal.CysAspGluGlyThrIleLeuCysAspValIl 67

2438 AGTTCAAATCCATGGAGGAGTGTTCATCCTAGAAAACTCCATGCAAG 2389
     |||||::::::| :::  ::|||        ||| |||||| :::|||
  67 eCysGluAsp...IleGlyAspCysProAsnProGluIleProMetGlyG 83

2388 AGCTACATTAAACGAAGCTGCAGGTTAAGGGCTTAGAGATGGGAAACCA 2339
     ||      |||  ||||     |||       ||| ||| |||:|||
  83 luCys......CysProValCysGlyGluGlyGlnTyrGlnThr 95

2338 GGTGACTGAGTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTGTCTCA 2289
     ||||||    ::::||| |||||||||||:|||||
  96 GlySer..ValValGluGlyProLysGlyGluThr.......... 106

2288 ACTAGGAGGCTAGCTGTTAACCCTGAGCCTGGGTAATCCACCTGCAGAGT 2239
     :::::|||    :::|||    :::::|||:::|||
 107 ........................GlyProArgGlyGluA 112

2238 CCCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCGTATAAGTCCAG 2189
     :::|||    :::  ::::|||        ::::::|||
```

```
 112 rgGlyProProGlyAlaProGlyArgGlyAspGlyIleProGly..........  125

2188 ACTGAAACCCCCTTGGAGACGGCCTCCAGTCAGGCAGCCCTAGAGAGACTGGGG 2139
                               |||  |||  |||
 126 ..............................GlnProGlyIlePro...........  130

2138 AGAGAGGAGGGACGCGCCCCCAGCCCCAGCTGTCCAGCTACGCACCACCTCAG 2089
             |||   |||||| ||||||::
 131 ..............GlyProProGlyProProGly....................ProA  139

2088 CAGCACAGGGTGGCGCAGAGAGCCACATTACTTTGGCAGCAACAGAAAC 2039
     ||::: ||||::       ::: :::||||||
 139 laGlyLeuGlyGlyAsn.................PheAlaProGlnMetSer  150

2038 TGGCGGCCAGCCCGGCAGCCCCATGGGCTAACAGGAGGCGGGGAGCGTGGG 1989
     :::                 :::            ||  :::|||
 151 TyrGlyTyrAspGluLysSerAlaGlyGlyIleSerMetProGlyProMetGl  167

1988 ACCCAGTGAGGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTTCTACGCT 1939
     ||||||               ||||||::::::|:: |||||
 167 yProMetGlyProArgGlyProProGlyProSerGlySer..........  180

1938 GAGTATTTGGCCAAGTCGCTCTTGTCAAATACTACCTGTGTAGCAAAGTA 1889
     ||||||
 180 .......................................................  180

1888 AATGGCGACCAGACCCCAGGCTCGCGGAGACACCATATAGGCAGTGACAG 1839
     |||   |||:::                ||| :::
 181 ProGly..ProGln...GlyPheGlnGlyProGlyGluProGlyGlu  195

1838 ACTGGCTGAGCTGGACAATGAGCGCCCATAAACAGGGATGGGCCCACCTGG 1789
     ||| |||||             ::: |||        |||||||||||
 196 ProGlyAlaAlaGlyAlaMetGlyPro.......ArgGlyProProGl  209

1788 GACAGCAGGAAGGCAC.............................TATC 1769
     |                                          |
 209 yProProGlyLysAsnGlyGluGluAspGlyGlyGluAlaGlyLysProGlyArgP  226

1768 CAGGATGGCGAGTGCCAGGCAGATGCCCCGGGCCCGGAA........... 1731
     ||||| |||||                ||||||:::
 226 roGlyGluArgGlyPro..........ProGlyProGlnGlyAlaArgGly  239

1730 ..........CCACCCTGGCCTCGGTGGGCTCACCCACCACCACGTAC. 1691
     ||||||               |||:::  :::
 240 LeuProGlyThr.AlaGlyLeuProGlyMetLysGlyHisArgGlyPheA  256

1690 ..........GGAGACAATCACAGGCAGGAGGCCCCGCCAGAG 1661
     |||||||            |||  |||:::::
 256 snGlyLeuAspGlyAlaLysGlyAspThrGlyProAlaGly.ProLysGl  272

1660 CGCGGGTGCGAGGT........................GGGAGCA 1641
     :|||||||                              |||
 272 yGluProGlyAsnProGlyAsnGlyAlaProGlyGlnValGlyProA  289

1640 GGCCACTGCCT..............CCAGCACCACGTGTCCATTAGGG 1606
     ||    |||||              |||::::|||  |||
 289 rgGlyLeuProGlyGluArgGlyArgGlyProSerGlyProAlaGlyGly  305

1605 ..........AAGGGAGCTCCAGGCTTAGGGCCTGGCAGGAAGCTGGT 1568
     |||:::||||||              :::
 306 AlaArgGlyAsnAspGlyThrProGlyAlaAla...................  316

1567 CATCAGGCTGTCCTCACTGCTAGCACCTCCAGTGTCCCCTCGGTATTTGG 1518
     ::::|||||              |||:::
 317 ................................GlyProPro.GlyProThrGlyPro...  324

1517 GCAGGAACACCTGCTTCTCCCGGTGGTAGAGGGAGGCCAGTGTGTAGGGC 1468
     ::|||   |||:::    |||||||||    ||| |||
 325 ThrGlyProGlyPheProGlyProGlyValGly...ProLysGlyAspAl  340

1467 AGGATCTGCAGGGCTGAGAAGGTGAACCCGGTGAGGGCGGCTGAAGCTGT 1418
     |||  |||             |||
 340 aGlyProGlyGlnGlySerArg.......................G  347
```

```
1417 CACCACGGCCCACACTGTGGACAGCAGGCATGTGGCACCGGCAGCCACAGGGA 1358
     ::|||                                         ||||||||||
 347 lyPro..............................................AspGlyProGlnGly  353

1367 AAGCTG.....CCACACTGGCCAAATAGACTGCTCGAGTGCCGAATCGG 1324
     |||            |||::::::::
 354 AlaArgGlyGluProGlyAlaProGlyGln..............................Al  364

1323 TGCACCAGCCGGT..........CCAATGACCAGAGAGAAGACCAGGGA 1286
     |:::|||||||                 ::::|||||||
 364 aGlyProAlaGlySerAlaGlyAsnProGlyThrAspGlyGlnProGlyA  381

1285 GATGGCCGCCACTGCAGGAACA...............GCCCCAGGCTGCCCA 1251
     :::                                    |||||||||  |||
 381 lalysGlyAlaThrGlyAlaProGlyLylleAlaGlyAlaProGlyPhePro  397

1250 TCCGAACGCCTTCATCATAGTGTCTCCGGCCTCGGTGCCGGCTCAGCT 1201
     ::: ||             ||||||:::: |||||
 398 GlyAlaArg...........GlyAlaProGlyGln..............  406

1200 CTGGGCACGCCTGGTACAGCCCCTCGCCCACGAAATCCGTGTAAAACAG 1151
     ::: |||                :::|||::::|:
 407 ......GlyProGlyGlySerProGlyProLysGlyAsn...........  417

1150 CGTGAAGTCATGAGTGCCATCCAGCTGCACAGCTCAGCACCAAGAAGACCC 1101
     :::::::|:|||                  :::||||
 418 ....................................AsnGlyGluPro......  421

1100 GGGCCAGGGTGCCGGGCATGCGGCAGCACAGCT.....GGTGCAGCCGG 1057
     |||||   |||:::  ::                     |||||| ||
 422 ..........GlyAlaGlnGlyAsnLysGlyGluProGlyGlyAlaLysGl  434

1056 GGAAGCAGGGCGCCCAGGTTCC...GGAAAGCCAAGCGGCGGCCCGGGATGG 1010
     |||::|||              :::                 |||||||||
 434 yGluSerGlyProAlaGlySerGlnGlyProProGlyProProGlyGluG  451

1009 ACAGCAGTGGGGCGACAAGGAGGGGGCCGACAGCCCCTTCGCTCGGCTCGG 960
     :::::::|:::||||                          |||
 451 luGlyLysArgGlySerArgGlyGluPro....................  460

 959 TGGGGCCCAGTCAATGCGGCTGCCTCCAGCCAGCAGTGTGGCTGCTACGCAG 910
     |||||::    |||||||||
 461 ..GlyProSerGlyProProGlyPro.......................  468

 909 GTGAGGAAGATGAGGGTGAGCAGGCCAAAGAGGCCACTCCT........... 870
     ::|||||||||||||::::||||
 469 ......................AlaGlyGluArgGlyAlaProGlySerArgG  479

 869 ......CCTGGGTGCCCAGGTAGGGGG..........CCAGGGCAC 840
     ||||||||::::        |||                ||||:::
 479 lyPheProGlyAlaAspGlyAlaGlyLyProLysGlyProGlyGlu  495

 839 TGGTGTCCCAGTCAATGCAGGCAGGAGGTAGCCCAGGCAGCCCCAAGA 790
     |||                 ::::::|||||        |||||||::
 496 ArgGlyProVal.....GlySerAlaGlyProLysGlySerProGlyGl  510

 789 CTGATCATGAAGGCATAGACAGAGTAGGCCTTGGCGACAGTGGT.... 747
     |||||||||||::|||||||
 510 u................................SerGlyArgProGlyGluProGlyLeuProG  521

 746 ...............................................CCGGGTCCC....GGAAGA 732
     ::::|||||             |||:::
 521 lyAlaLysGlyLeuThrGlySerProGlySerProGlySerProGlyLys  537

 731 GGTCAGAGAGCAGGGCCTCCAGTGAGTGAAGCACACCTGGCCACAGAAG 682
     ::|||   ::|||             ::: |||  ||||||:
 538 ThrGlyProAlaGlyAlaProGlyGlnAspGlyAlaArgGlyProProGl  554

 681 TCCCAGCAGCCCCACGC..................CCAGGATGAG 656
     ::|| :: |||                        ||||||
 554 yProProGlyAlaArgGlyGlnSerGlyValMetGlyPheProGly....  569
```

```
RA   SHAKYA R., STURM A.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y16768; CAA76369.1; -.
DR   PFAM; PF00083; sugar_tr; 1.
SQ   SEQUENCE  515 AA;  54424 MW;  9F3E27DF CRC32;

alignment_scores:
            Quality:  347.50                    Length:     505
              Ratio:    1.292                      Gaps:      17
  Percent Similarity:   53.267    Percent Identity:    25.941

alignment_block:
US-09-030-606-110 x 065803

Align seg 1/1  to: 065803   from: 1 to: 515
```

```
332 CAGCTCTTGCTGGTCAACCTGCTAACCTTGGCCTGGAGGTGTGTTTGGC 381
    :::||||:|||::::|   ::||||:::|||  :::  ::  :::
 34 LysLeuValLeuValAlaAlaAlaIleAlaAlaGlyValGlnPheGlyTrpAl  50

382 CGGCAGGCATCACCTATGTGCCCGCCTCGTGCTGCTGGAAGTGGGGTAGAGG 431
    |  :::::: ::: :::    |||    :::::::  :::|||:| :
 50 aLeuGlnLeuSerLeuLeuThrProTyrValGlnLeuLeuLeuGlyIleProH  67

432 AGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCCTGGGCCTGGTC 481
    :||:||  |||:::   :: :||||||  |||||||  |||:::::  :
 67 isLysTrpAlaAlaTyrIleTrpLeuCysGlyProIleSerGlyMetLeu  83

482 TGTGTCCGGCTCCTAGGCTCAGCCAGTGACCACTGGCGCTGGACGCCATGG 531
    :: :::::|||||||  :::::|||||  ::::::::
 84 ValGlnProIleValGlyTyrTyrSerAspHisCysGlnSerSerPheGl 100

532 CCGCCGCCGGCCGTTCAGGCACTCCA  GAACTGCTTCGTCTCCGGCTCTGCT 581
    |||||||||||||||||  |||  :::|||:::::::|||:::
100 yArgArgArgProPheIleAlaSerGlyAlaGlyCysValAlaIleSerV 117

582 TCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTG............ 619
    ::::|||||  |||   |||
117 alIleLeuIle.........GlyPheAlaAlaAlaSerIleSerTyrLysAla 130

620 ........CTGTGCCCGGATCCCAGGCCCCTGGAGCTGGCACTGGCTCAT 660
    ::::||| :::::::::|||:::::::
131 GlyAspAspMetSerLysThrLeuLysProArgAlaValThrValPheVa 147

661 CCTGGGCCTGGGGCTTCTGGCACCTGTGGCCCAGGTGTGTCCACTCCAC 710
    :::::||  |||||||:::  ::: |||     :::||| :::::|||
147 lIleGlyPheIleLeuAspValAlaAsnAsnMetLeuGlnLeuProC 164

711 TGGAGGGCCCTGCTCTGACCCTTCCGG...GACCCGGACCACTGTCGC 757
    ::||||||||||:::|||||  :: : ||| |||
164 ysArgAlaLeuLeuLeuAlaAspLeuCysSerGlyAspThrArgArgMetArg 180

758 CAGGCTACTCTGTCATCATGCCTTCATGATCAGTCTTGGGGGCTGCCTGGG 807
    :::|||   ::||||||||||||:::||:::::||||| :||:
181 SerAlaAsnAlaPheTyrSerPheMetAlaValGlyAsnIleLeuGl 197

808 CTAC............................CTCCTGCCTGCCATTG 827
    |||
197 yTyrAlaAlaGlySerTyrAsnAsnLeuTyrLysLeuProPheSer 213

828 ACTGGGACACCAGTGCCCTGGCCCCTACCTGGCACCCAGGAGGAGTGC 877
    |||   |||   |||||||||  :::  :::::||
214 .....LysThrHisAlaCysAspLeuTyrCysAlaAsnLeuLysSerCys 228

878 CTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCTAGCAGCCACACT 927
    :::::|||||||:::::|||||||::::|  |||::||
229 PheIleIleSerIleAlaLeuLeuIleLeuIleThrValValAlaLeuSe 245

928 GCTGGTGGCTGGAGGGAGCAGCCGCTGGGCCCCACCGAGCCAGCAGAGGGC 977
    ::::|| ||||||  |||||  ::::::|||||
245 rValValArgGluAsnSer......GlyProProAspAlaAspAlaA 260
```

```
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PRO ALPHA 1(I) COLLAGEN.
GN   COL1A1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85130970.
RA   CHU M.L., DE WET W., BERNARD M., RAMIREZ F.;
RT   "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT   Promoter structure, AluI repeats, and polymorphic transcripts.";
RL   J. Biol. Chem. 260:2315-2320(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88329734.
RA   D'ALESSIO M., BERNARD M., PRETORIUS P.J., DE WET W., RAMIREZ F.;
RT   "Complete nucleotide sequence of the region encompassing the first
RT   twenty-five exons of the human pro alpha 1(I) collagen gene
RT   (COL1A1).";
RL   Gene 67:105-115(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89025644.
RA   TROMP G., KUIVANIEMI H., STACEY A., SHIKATA H., BALDWIN C.T.,
RA   JAENISCH R., PROCKOP D.J.;
RT   "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT   chain of human type I procollagen.";
RL   Biochem. J. 253:919-922(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91138770.
RA   MAARTA A., BORNSTEIN P., PENTTINEN R.P.;
RT   "Highly conserved sequences in the 3'-untranslated region of the
RT   COL1A1 gene bind cell-specific nuclear proteins.";
RL   FEBS Lett. 279:9-13(1991).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92157916.
RA   WESTERHAUSEN A., CONSTANTINOU C.D., PACK M., PENG M.Z., HANNING C.,
RA   OLSEN A.S., PROCKOP D.J.;
RT   "Completion of the last half of the structure of the human gene for
RT   the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL   Matrix 11:375-379(1991).
RN   [6]
RP   SEQUENCE FROM N.A.
RA   KORKKO J.M., EARLEY J.J., NUYTINCK L., DEPAEPE A., PROCKOP D.J.,
RA   ALA-KOKKO L.;
RT   "Analysis of the COL1A1 and COL1A2 Genes by CSGE and DNA Sequencing
RT   in 12 Patients with mild OI (Type I). Identification of Common
RT   Sequences for Null Allele Mutations.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF017178; AAB94054.1; -.
DR   PFAM; PF01410; COLFI; 1.
DR   PFAM; PF01391; Collagen; 18.
DR   PFAM; PF00093; vwc; 1.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE  1461 AA;  138672 MW;  4AD77546 CRC32;
```

```
 978 TGTCGGCCCCTCCTTGTGCGCCCACTGCTGTCCATGCCGGCGCCCGCTTG 1027
     ::: |||       |||             :::::||| |||
 260 laGluGluPro...........ProSerGlyLysIle 269

1028 GCTTTCGGAACTGGGCGCCGCCCTGCTTCCCCGGCTGCACCAGCTGTGCTG 1077
     ::: |||   |||||||       ||| ::|||
 270 ProVal......PheGlyGluLeuGlyAlaLeuLysAspLeu..... 282

1078 CCGGATGCCCCGCACCCTGCGCCGGCGCTCTTCGTGCTGCTGAGCGTGTGCAGCT 1127
     ||||||  ::      |||::|||||        :::|
 283 ......ProArgProMetLeuLeuLeuLeuIleValThrCysLeuAsnT 297

1128 GGATGGCACTCATGACCTTCACGCGTGTTTTACACGGATTTCGTGGGCGAG 1177
     ||:::||  :: |||  ||||||  |||||||::||::|||::
 297 rPIleAlaTrpPheProPheIleLeuPheAspThrAspTrpMetGlyArg 313

1178 GGGCTGTACCAGGGCGGTGCCCAGAGCTGAGCGGGCACCCAGGGCCCGGAG 1227
     :::::::|| |||:::      |||      ||| ||   :::
 314 GluIleTyrGlnGlyGlyThr.....AlaGlyGlnGly.........Ly 324

1228 ACACTATGATGAAGGCGTTCGGATGGCAGCCTGGGCGTGTTCCTGCAGT 1277
     : ||||||||||||||| ||::|||||||||||||::||:::
 324 sLeuTyrAspGlnGlyValArgAlaGlyValLeuGlyLeuLeuLeuAsnS 341

1278 GCGCCATCTCCCCTGCTTCTCTCGGTCATGACCGGCTGGTGCAGCGA 1327
     ::::: |||  :: :|||:: |||       ||||||::
 341 erValValLeuGlyLeuThrSerIleAlaValGluTyrLeuValArgGly 357

1328 TTCGGCACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTGTGGC 1377
     :::  |||  ||| :::  :: |||  || :::|
 358 ValGlyGlyValLysIle...LeuTrpGlyPheValAsnPheIleLeuAL 373

1378 TGCCGGTGCCACATGCCTG............................ 1396
     |     :: ||:::
 373 aIleGlyLeuValMetThrValValSerLysValAlaGlnHisGlnA 390

1397 ..TCCCACAGTGTG.......................GCCGTGGTGACA 1420
     ::: ||||||:       ||  |||
 390 rgGluHisSerAlaAsnGlyGlnLeuLeuProSerAlaGlyValLys 406

1421 GCTTCAGCC............GCCCTCACCGGTTCACCTTCTCAGCCCT 1458
     ||::::||||    |||::::     ||::|
 407 AlaGlyAlaLeuSerLeuPheSerIleLeuGlyIleLeuProLeuSerIleTh 423

1459 GCAGATCCTGCCCTACACACTGGCTCCCCTCTACCACCGGAGAAGCAGG 1508
     :::||||||||||||||::||         :::
 423 rTyrSerIleProPheAlaLeuAlaSerIleTyrSerSerGlySerGlyA 440

1509 TGTTCCTGCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGAC 1558
     :::||       ::    |||      ||| |||
 440 laGlyGlnGlyLeuSerLeuGlyValLeuAsnLeuAlaIleValAlaVaIPro 456

1559 AGCCTGATGACCAGCTTCCTGCCAGGCGCCTAAGCCTGGA.......... 1597
     ::::|| |||  :||||   :::
 457 GlnMetIleValSerValLeuValAlaGlyProPheAspSerLeuPheGlyGl 473

1598 .........GCTCCCCTTCCCTAATGGACACGTGGGTGCTGAGGCAGTG 1637
     ::::||||::|||  ||||||:::||||
 473 yGlyAsnLeuProAlaPheValValGlyAlaIleSerAlaAlaIleSerG 490

1638 GCCTGCTCCCA............CCTCCACCCCGGCTGCGGGGCCTCT 1675
     ||:::|||       |||   ||||||::|||||||
 490 lyValLeuAlaIleLeuValLeuProLysProSerLysAspAlaAlaSer 506
```

```
1676 GCCTGTGATGTCC 1690
     ::::: |||
 507 LysLeuSerLeuSer 511
```

seq_name: sp_human.O76045

seq_documentation_block:
```
ID   O76045     PRELIMINARY;     PRT;   1461 AA.
AC   O76045;
```

alignment_scores:
```
    Quality:        343.50          Length:        1061
    Ratio:            0.795          Gaps:            70
Percent Similarity:   40.716   Percent Identity:   26.673
```

alignment_block:
US-09-030-606-110/rev x O76045

Align seg 1/1 to: O76045 from: 1 to: 1461

2591 CTGTGTCTCGTGATGGCAACAGA.................... 2558

```
  60  IleCysValCysAspAsnGlyLysValLeuCysAspAspValIleCysAs  76          291  .....GluAsnGlyAlaProGlyGln..MetGlyProArgGlyLeuPr  304
      :::!!!:.!!!!!!:::::::                                                   :::!!                    !!!!!!:::!!!
2567 .AGGACCAACAGGCCACATCCTGATAAAAGGTAAGAGGGGGTGATCAG 2519       1791 TGGGACAGGAGGAGGCACTATCCAGGATGGCGAGGTCCAGGCAGATGCC 1742
      ::!!!                                                                  !!!!!!         !!!!!!:::!!!
  76  pGluThr...........................                78          304  oGlyGluArgGlyArg.......ProGlyAlaProGlyProAlaGlyAlaA 319

2518 CAAAAAGACAGTGCTGTGGGCTGAGGGGACCT..GGTTCTTGTGTGT.. 2474       1741 CCGGCCCGGAA.............................CACCCTGGCCTCGGTGG 1713
      ::!!:::!!! !!! ::::!!! !!!::!!! !!!                                   !!! !!!!!!:::!!!
  79  .....LysAsnCysProGlyAlaGluvalProGluGlyGluCys.CysPr  93        319  rgGlyAsnAspGlyAlaThrGlyAlaAlaAlaGlyLProGlyProThrGly 335

2473 ....TGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTCAAATC 2428       1712 GCTCACCCACCACCACA.......................CGTACGGAGAC 1685
      !!!!!!                                                                 !!!!!!              ::: !!!!!!
  93  ovalCysProAspGly.....................             98          336  ProAlaGlyProProGlyPheProGlyAlaValGlyAlaLysGlyGluAl 352

2427 CCATGGAGGAGTGTTCATCCTAGAAACTCCCATGCAAGAGCTACATTAA 2378       1684 ATCACAG................GAGAGGCCCCAGAGCGCGGGTGGAG 1650
      !!!::!!!                                                               ::::!!    !!!!!!:::!!!:::!!!!!!
  99  ......................SerGluSerPro...............                102          352  aGlyProGlnGlyProArgGlySerGluGlyProGlnGlyValArgGly. 368

2377 ACGAAGCTCAGGTAAGGGGCTT............................ 2354       1649 GTGGGAGCAGGCCACTGCCTCCAGCACCCACGTGTCCATTAGG...AAG 1603
      !!!   ::::!!! !!!::                                                    :::!!! !!!!!!:::!!!!!!:::!!!!!!
 103  ThrAspGluThrThrGlyValGluGlyAspThrGlyProArgGlyPr  119         369  ........GluProGlyProGlyProAlaGlyAlaAlaGlyProAla  382

2353 ........................AGAGATGGGAAAACCAGGTGACTGAG 2329       1602 GGAGTCCAGGCTTAGGG......CCTGGCCAGGAAGCTGGTCATCAGGCT 1559
      !!!!!! !!!!!!:::                                                       !!! !!!!!!   !!! ::::!!!:::!!!
 119  oArgGlyProAlaGlyProProGlyArgAspGlyIleProGlyGlnProG 136        383  GlyAsnProGlyAlaAspGlyGlnProGlyAlaLys.GlyAlaAsnGlyA 399

2328 TTTATTCAGCTCCCAAAAACCCT.....TCTCTAGGTGTCTCCAACTA 2285       1558 GTCCTCACTGCTAGCACACCTCCAGTGTCCCTCGGTATTGGGCAGGAACA 1509
      ::!!! ::!!!                     !!!!!!:::                              !!!!!!    !!!!!! !!!!!!
 136  lyLeuProGlyProProGlyProProGlyProProGlyProGlyLeu 152          399  laPro.............GlyIle..AlaGlyAla  405

2284 GGAGGCTAGCTGTA...ACCCTGGCCTGGTAATCCACCTGCAGAGTC 2238       1508 CCTGCTTCTCCGGTGGTAGGAGGAGGCCAGTGTGTAGGCAGGATCTGC 1459
      !!!!!! :::   !!! :::!!!:::!!!!!!                                       !!!:::  !!!!!!:::!!!  :::!!!
 153  GlyGlyAsnPheAlaProGlnLeuSerTyrGlyTyrAspGluLysSerTh 169        406  ProGlyPheProGlyAlaArgGly..........ProSerGlyPro.G 418

2237 CCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGTCCAGA 2188       1458 AGGGCTGAGAGGTGAACCCGGTGAGGCGCGTGAAGCTGTCACCACGGC 1409
      !!!!!! :::!!! !!!                                                      !!!!!! !!! !!!!!!:::  :::!!!
 169  rGlyGlyIleSerValProGlyProMetGly.............      179          418  lnGlyProGlyGlyProProGlyProLysGlyAsnSerGlyGluProGly 434

2187 CTGAAACCCCCTTGGAAGGCCTCCAGTCAGGCCTCAGCCCTAGAGACTGGGGA 2138       1408 CACACTGTGGGACAGGCATGGCACCCGGCAGCCACAGGGAAAGCTGCCA 1359
      !!!!!!                                                                 :::  !!!:::!!!!!!   :::!!!
 180  ........................ProSerGlyPro...........       183          435  ...AlaProGlySerLysGlyAspThrGly.AlaLysGlyGlyProGlyP 450

2137 GAGAGGAGGAGCGACCCCCAGCCGTGCAGCTACGCCACTCAGC 2088       1358 CACTGCCAAATAGACTGCTCGAGTGCCGAATCGTCGCCAGCCGGTCC 1309
      !!!!!! !!! !!!!!!:::::!!!                                              !!!!!!:::!!!   :::!!! !!!!!!
 184  .......ArgGlyLeuProGlyProProGlyAlaPro...GlyProGlnGl 197        450  ro.................Val.GlyValGlnGlyProProGlyPr 459

2087 AGCACAGGTGGCAGCAGAGAGCCACATTACTTTGGCAGCAACAGAAACT 2038       1308 ATGACCAGAG.......AGAAGACCAGGAGATGGGCCACTGCAGGAACAG 1265
      :::!!!                                                               ! :::!!!                 !!!  :::!!!
 197  yPheGlnGlyProProGlyProGlyGluProGlyProGlyAlaSerGlyProM 214        459  oAlaGlyGluGluGlyLysArgGlyAlaArgGlyGluGlyProThrG 476

2037 GGGCGCCAGCCCGGCCAGCCCATGGGAGCTAACAGGAGCGGGAGCTGGA 1988       1264 CC...CCAGGCTGCCCATCCGAACGC...................... 1242
      :::!!!                                                                 CC !!!!!! !!!::!!!!!!
 214  etGlyProArgGlyProProGlyProProGlyLysAsnGlyAspAspGly 230        476  lyLeuProGlyProGlyGluArgGlyGlyProGlySerArgGlyPhe 492

1987 ......................CCCAGTGAGGCAGGCCCTCCA.... 1967       1241 ...CTTCATCATAGTGTCTCCGGCCTCCGGTGCCGGCTCAGCTCTGGGC 1195
      !!!!!! !!!!!!                                                          :::!!!                 :::!!! !!!!!!
 231  GluAlaGlyLysProGlyArgProGlyGluArgGlyProProGlyProGl 247        493  ProGlyAlaAspGlyValAlaGlyProLysGlyProAlaGlyLeuArgGl 509

1966 ........................CCCCAATGTGCTGGAAGTTTCTACGCTGAGTAT 1933       1194 ACGCCCTGGTACAGCCCCTCCGCCACGAAATCCGTGTAAAACAGCGTGAA 1145
      !!!!!! :::!!!                                                          !!!!!!             :::!!!
 247  .............LeuP 259                                            509  ySerProGlyProAlaGlyProLysGlySerPro............ 520

1932 TTGGCCAAGTGCTCTGTCAAATACTACCTGTGTAAGAAGTAAATGGC 1883       1144 GGTCATGAGTGCCATCCAGCTGCACAGTCTCGCAGAGAGCCGGCCGCA 1095
      :::!!! :::    !!!                                                      ::: !!!!!! !!!!!! !!!!!!
 259  nGlyAlaArgGlyLeuProGlyThrAlaGly.............      275          520  ...........................GlyAlaGlyArgProGlyGluAlaGlyLyLe 531

1882 GACCAGACCCAGGCTGCGGCAGACCACCATATAGGCAGTGACAGACTGGC 1833       1094 GGGTGCGGGGCATGCGCAGCAGCAGCCTGGTGCAGCCGGGGAAGCAGGGCG 1045
      !!!                                                                                          !!!!!!
 276  AspAlaGlyProAlaGlyProLysGlyGluProGlySerProGly...... 290       1044 CCCAGGT.............TCCGAAAGCCAAGCGGCCCCGGCATGGACA 1007

1832 TGAGCTGGACAATGGAGCCCATAAACAGGGATGGGGCCA..........CC 1792       531  uProGlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAspG 548
      ::!!!!!!!!!                !!! :!!!!!!
```

```
373 CACCTCCAGGC.........CAAAGGTTAGCAGGTTGACCAGCAGA 336
      |||||||||              |||  ||||::
813 ..ProProGlyAlaAspGlnProGlyAlaLysGlyGluProGlyAsp 828

335 GCTGGGCTTCCGGTGCCGCAGCAGGCGGCTCACCCAGCCTCTGGACC 286
      |||||  :::||:|||  |||  |||   :|:||||
829 AlaGlyAlaLysGlyAspAla......GlyProProGlyAlaGlyPr 843

285 ATAGTGGGCCA.........GGCGGGTAGGGCTCAGGGGGCCGTTCA 248
      |  :::||||||           |||  |||  |||    :::
843 o.AlaGlyProProGlyProIleGlyAsnValGlyAlaProGlyAlaLys 859

247 GGCACTCCAGAACTGCTTCGTCTCGGCTCTGCT......CCAAAGCTGC 204
      |||:::        |||||  |||  |||         |||  |||:::
860 GlyAla......ArgGlySerAlaGlyProProGlyAlaTh 871

203 GGCCTCCTCCCTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCC 154
      :::||||                :::             |||:::|||::
871 rGlyPheProGlyAlaAlaGlyAlaArgValGlyProProGlySerGlyA 888

153 ATTTCTGCCAGCCCTTTGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAAC 104
      ::  |||           |||||:::|||       ||||||||::::
888 snAlaGlyProProGlyProAlaGlyLysGlu.......... 901

103 ACCTGCTGCTGTGG......GCACCTCAGTGGGACACGTCTCATCACT 60
      |||           :||  |||||  :||  |||  ::
902 ..........GlyGlyLysGlyProArg.GlyGluThrGlyProAlaG 914

59 CAGATCCTGGCCCA........GGCGGCG 40
     ::  |||||:::
914 lyArgProGlyGluValGlyProProGlyProProGlyAlaGlyGlu 930

39 CGGCTGTCACCCGGGAGCC 22
     :::  |||||||||||
931 LysGlySerProGlyAla 936
```

seq_name: sp_plant:O04077

seq_documentation_block:

alignment_scores:
```
    Quality:       337.50               Length:       520
      Ratio:         1.255                 Gaps:          20
Percent Similarity:   51.731       Percent Identity:  25.769
```

alignment_block:
US-09-030-606-110 x O04077

Align seg 1/1 to: O04077   from: 1 to: 523

```
1006 GCAGTGGGGCGACAAGAGAGGGGCCGACAGCCCTTCTGCTGGCTCGGTGG 957
       ::        |||||
548 ly.........LysThrGlyProPro................. 553

956 GGCCCAGCG.............CTGCCTCCTCAGCCACCA...... 930
       ||||||||               ||||||  ||||||
554 GlyProAlaGlyGlnAspGlyArgProGlyProProGlyProGlyAl 570

929 .........................GCAGTGT 923
                                |||:::
570 aArgGlyGlnAlaGlyValMetGlyPheProGlyProLysGlyAlaAla. 586

922 GGCTGCTACGCAGGTGAGGAAGATGAGGTGACGAGGCCAAAGAGGCACT 873
       |||:::    |||  ||||||::::|||||||:::
587 .........GlyGluProGlyLysAlaGlyAlaGlyGluArgGlyVal 597

872 .........CCTCCTGGGTGCCCAGT........ 855
       |||||||||||||
598 ProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGluAlaGl 614

854 ..................AGGGGGCCCAGGGCACTGGGTCCCCAGTC 827
                          |||||  :::
614 yAlaGlnGlyProProGlyProProGlyProAlaGlyProAlaGlyGluG 631

826 AATGGCAGGCAGGA..............GGTAGCCCAGGC 801
       ||                           |||
631 lnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyProAlaGly 647

800 AGCCCCAAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCGACAG 751
       |||           :::|||||
648 ProProGlyGluAlaGlyLysLysProGlyGluGlnGlyValProGlyAspLe 664

750 TGGT...CCGGGTCCC.......GGAAGAGGTCAGAGAGCAGGGCCTCCAG 710
       |||      |||||||||:::|||  |||
664 uGlyAlaProGlyProSerGlyArgGlyValGlyArg...GlyPheProG 680

709 TGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGCCCAGGA 660
       |||  :::||   |||              ::|||||  :::
680 lyGluArgGlyValGlnGlyLyProProGlyProAlaGlyProArgGlyAla 696

659 TGAGCAGTGCCA............................GCTCCAGG 640
       ::  |||  :|||                          |||||||
697 AsnGlyAlaProGlyAsnAspGlyAlaAlaLysGlyAlaAlaGlyAlaProGl 713

639 GGCCTGGGATCCGGGCACAGCAGCCCGTCGTAGCCTAGCCAGCCGGCCCTTGGGAT 590
       ||||       ::         :::
713 yAla......ProGlySerGlnGlyAlaProGlyLeuGlnGlyLeuMetProG 728

589 GAGAAAGAGGTCAGCAGGATGCCCAAGGACAGTGCCCAGATGAAGGGCC 540
       ||||||||:::|||||  |||  :::
728 lyGluArgGlyAlaAlaAlaGlyLeuProGly.....ProLysGlyAspArg 742

539 GGCGGGCCGCCATAGCGTCCACGCCCAGTGGTCACTGGCTAGGAGC 490
       ||||||||::::|||              :::
743 GlyAspProAlaGlyProLysGlyAlaAspGly........SerProGlyLy 756

489 G...GGACACAGA.......CCAGGCCCAGCAC 467
       ||  |||               ||||||||||||
756 sAspGlyValArgGlyLeuThrGlyLeuProIleGlyProGlyProGlyProAla. 772

466 TGGACCAATGCCCA..................... 453
       ::|||
773 .....GlyAlaProGlyAspLysGlyGluSerGlyProProAla 787

452 ..........GCACCATGGTCATGAACTTCTCCTGTACC 424
       :||  :||  |||||
788 GlyProThrGlyAlaArgGlyAlaProGlyAspArgGlyGluProGlyPr 804

423 CCCACTTCCAGCAGCAGCAGGCGGCACATAGGTGATGCCTGCGGCCAAACA 374
       |||  |||||||:::  |||:::
804 oProGlyProAlaGlyPheAlaGly.................. 812
```

332 CAGCTCTTGCTGGTGTCAACCTTGCCTGCTAACCTTTGGCCTGGAGGTGTGTTTGGC 381

```
 312 AspTrpMetGlyLysGluValTyr.................GlyGl 321
1213 CACCGAGGCCCGAGACAC...TATGATGAAGGCGTTCGGATGGGCAGCC 1259
     ||||                :::  |||||:::||||||  ||||:||
 321 yThrValGlyGluGlyHisAlaAlaTyrAspMetGlyValArgGluGlyAlaL 338
1260 TGGGGCTGTTCCTGCAGTGGCGGCCATCCCCTGGTCTTCTCTCCTGGTCATG 1309
     ||||||||| :::|||:: :::::  |:::  ||  |||||| |||||
 338 euGlyLeuMetLeuAsnSerValValLeuGlyAlaThrSerLeuGlyVal 354
1310 GACCGGCTGGTCGCAGCAGTTCGGCACTCGAGCAGTCTATTCGCCAGTGT 1359
     |||  |||:|||| |||     |||      |||    ::::|||
 355 AspIleLeuAlaArgGlyValGly...GlyValLysArgLeuTrpGlyIl 370
1360 GGCAGCTTTCCCTGTGGCTGCCGCGTGCCACATGCCGTCCCACAGTGTGG 1409
     :::|||               |||        |||  ::::|||
 370 evalAsnPheLeuLeu.........AlaIleCysLeuGlyLeuThrValL 384
1410 CCGTGGTG.......................................... 1417
     |||:||||
 384 euValThrLysLeuAlaGlnHisHisSerArgGlnTyrAlaProGlyThrGly 400
1418 ..............................ACAGCTTCAGCCGCCCT 1434
                                    ::::||||:::|||||
 401 AlaLeuGlyAspProLeuProProSerGluGlyIleLeuLysAlaGlyAlaLe 417
1435 CACCGGGTTCACCTTCTCAGCCTGCAGATC.................CTGC 1469
     ||||||||:::|||||              ::
 417 uThrLeuPheSerValLeuGlyLeuValProLeuAlaIleThrTyrSerIleP 434
1470 CCTACCACTGGCCTCCCTCCTACCACCGGAGAAGCAGGTGTTCCTGCCC 1519
     ||:::||||||||||||||||||||||||||:::|||
 434 roPheAlaLeuAlaSerIlePhe. 441
1520 AAATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGACAGGCTG...... 1564
     :::|||:::|||||||:::||||::
 442 .........SerSerThrSerGlyAlaGlyGlnGlyLeuSerLeuGlyVa 455
1565 ..........................ATGACCAGCTTCCTGCCCAG 1583
                              |||:::|||
 455 lLeuAsnLeuAlaIleValIleIleProGlnMetPheValSerValLeuSerG 472
1584 GCCCTAAGCCTGA..............GCTCCCTTCCCTAAT 1612
     |||||              |||||:::||||
 472 lyProTrpAspAlaLeuProPheGlyGlyGlyAsnLeuProAlaPheValVal 488
1613 GGACACGTGGGTGCTGAGGCAGTGGC..............CTGCTCCC 1647
     ||| :::||     ::|||||
 489 GlyAlaValAlaAlaAlaLeuAlaSerGlyIleLeuLeuSerIleLeuLeuPr 505
1648 ACCTCCACCC 1657
     :||||||
 505 oSerProPro 508
```

```
  36 LysIleMetValValAlaAlaSerIleAlaAlaGlyValGlnPheGlyTrpAl 52
 382 CGCAGGCATCACCATGCCGCCCTCGCTGCTGGAAGTGGGGTAGAGG 431
     :::||||            |||            :::|||:::||||
  52 aLeuGlnLeuSerLeuThrProTyrValGlnLeuLeuGlyIleLeuHisH 69
 432 AGAAGTTCATGACCATGGGTCTGGGCATTGCTGCTCGGCCTGGTC 481
     :: |||  ::  |||||:: :::|||:: |||   |||
  69 isThrTrpAlaAlaTyrIleTrpLeuCysGlyProIleSerGlyMetLeu 85
 482 TGTGTCCCGCTCCTCAGCCAGTGACCACTGCGTGCGACGTATGG 531
     |||  |||:||| |||      ||||||||   |||  :::|||
  86 ValGlnProIleValGlyTyrHisSerAspArgCysThrSerArgPheGl 102
 532 CCGCCGCCGGCCCCTCATCTGGCCACTGCCATCCTGCTGAGCC 581
     ||||||||||||   |||  |||     :: |||  :::||||
 102 yArgArgProPheIleAlaAlaGlySerIleAlaValAlaIleAlaValAlaV 119
 582 TCTTTCTCATCCCAAGGGCCGGCTGGCTAGCAGGGCTG.......... 619
     :::|||||||            |||  :: |||
 119 alPheLeuIle........GlyTyrAlaAlaLaAspLeuGlyHisSerPhe 132
 620 ..........CTGTGCCGGATCCCAGGCCCCTGGAGCTGGCCACTGCTCAT 660
                ||              |||||||||||
 133 GlyAspSerLeuAspProGlnLysLysValArgProArgAlaIleGlyIleIlePheva 149
 661 CCTGGGCGGTGGGGGCTGCTGGTCCTGTGGCCAGTGTGCTGCTTCACTCCAC 710
     :::|||||||||||||||||        :::|||||||  ||| |||
 149 lValGlyPheTrpIleLeuAspValAlaAsnAsnMetLeuGlnGlyLysProC 166
 711 TGGAGCCCCTGCTCTCTGACCTCTTCCGG...GACCCGGACCACTGTCGC 757
     ::||||||||||||||:::||||||||         :: |||
 166 ysArgAlaLeuLeuGlyLeuAspLeuCysProLeuCysSerAlaGlyLysAsnGlnAsnArgLysThrArg 182
 758 CAGGCCTACTCTGTCTATGCCTCTATGATCAGTCTTGGGGGCTGCCTGGG 807
     :::|||            :::||||       ||||  :::|||||
 183 AsnAlaAlaAsnAlaPhePheSerPhePheMetMetAlaValGlyAsnValLeuG 199
 808 CTACCTCCTGCTGCCATTGAC.................TGGG 833
     |||||||||            |||              ||||
 199 yTyrAlaAlaGlyAlaTyrSerLysLeuTyrHisValPheProPheThrL 216
 834 ACACCAGTGCCCTGGCCCCTACCTGGCCACCAGGAGGAGTGCCTCTTT 883
     :: |||||||      |||  :::||||||||||:::||||
 216 ysThrLysAlaCysAsnValTyrCysAlaAsnLeuLysSerCysPhePhe 232
 884 GGCCTGCTCACCCTCATCTTCCTCACCTGGGTAGCAGCCCACACTGCTGGT 933
     :: |||  :::|||||||   |||     :::||||||
 233 ......LeuSerIleAlaLeuLeuThrLeuValLeuLeuAlaThrSerAlaLeuIle 247
 934 G...GCTGAGGAGGCAGCCAGTGGGGCCCCACCGAGCCAGCA.......... 970
     :   :::||||||  |||||||      |||
 247 eTyrValLysGluThrValGluThrProGluLysThrValThrValThrThrG 264
 971 ..GAAGGGCTGTCGCCCCTCCTGTCGCCCCCACTGCTGTCCATGCCGG 1018
       :::|||||            |||  :::||||||
 264 ...ProCysPhe 275
1019 GCCCGCTTG....GCTTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCT 1062
     :::|||       |||||||   |||||:::|||
 276 GlyGlnLeuSerGlyValAlaPheLysGluLeu... 285
1063 GCACCAGTGCTGTCGCCCCATGCCCCGGCCACCCTGCCGGCTCTTCGTGG 1112
     :: |||  :::|||||            :::|||||
 286 ..........LysArgProMetTrpIleLeuLeuLeuLeuLeuV 295
1113 CTGAGCTGTGCAGCTGGATGGCACTCATGACCTTCACGCTGTTTTACACG 1162
     :: :::|||  |||||||        ||| |||   :::|||
 295 alThrCysLeuAsnTrpIleAlaTrpPheProPheLeuAspAsnAspThr 311
1163 GATTTCGTGGGGCCGAGGGCCTGTACCAGGGCCGTGCCCAGAGCTGAGCCGGG 1212
     :::|||                   :: :::|||
```

seq_name: sp_vertebrate:Q92029

seq_documentation_block:
ID   Q92029   PRELIMINARY;   PRT;   886 AA.
AC   Q92029;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   ALPHA-1 COLLAGEN TYPE III (FRAGMENT).
GN   COL3A1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC   Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 9266842.
RA   NAH H.D., NIU Z., ADAMS S.L.;
RT   "An alternative transcript of the chick type III collagen gene that

```
RT    does not encode type III collagen.";
RL    J. Biol. Chem. 269:16443-16448(1994).
DR    EMBL; U07973; AAA83407.1; -
DR    PFAM; PF01391; Collagen; 11.
DR    PFAM; PF00093; VWC; 1.
DR    PROSITE; PS01208; VWFC; 1.
FT    NON_TER    886    886
SQ    SEQUENCE    886 AA;    81916 MW;    FE778D29 CRC32;

alignment_scores:
        Quality:    335.50         Length:    998
        Ratio:     0.826          Gaps:       65
Percent Similarity:  40.681    Percent Identity:  27.054

alignment_block:
US-09-030-606-110/rev x Q92029

Align seg 1/1    to: Q92029    from: 1    to: 886

2479 GTGTGT...TGCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTC 2433
          :::|||||||::
  51 IleCysValCysAspSerGlySerValLeuCysAspSerIleIleCysAs 67

2432 AAATCCCATGGAGGAGTGTTCATCCTAGAAACTCCCATGCAAGAGCTAC 2383
     ::|||        |||    |||    |
  67 pAspGlnGluLeuAspCysProAsnProGluIleProLeuGlyGluCys. 83

2382 ATTAAACGAAGCTGCAGGTTAAGGGGCTTAGAGATGGGAAACCAGGTGAC 2333
     |||          :::||    :::
  84 ........CysProValCysProGlnThrThrProGlnProThrGlu 96

2332 TGAGTTATTCA..GCTCCCAAAAACCCTTCTCTAGTGTGTCTCAACT 2286
     ||:::          |||||||||
  97 LeuProTyrThrGlnGlyProLys................... 104

2285 AGGAGGCTAGCTGTTAACCCTGGCCTGGTAATCCA.......CCTGCAG 2242
     ||||:::||    |||    |||
 105 ..............GlyAspProGlySerProGlyS 112

2241 AGTCCCCGCATTCCAGTGCATGGAGCCCTTCTGGCCTCCCTGTATAAGTC 2192
     ::|||:::|||    :::|||||
 112 erProGlyArgThrGlyAlaProGlyProGlyProGlnPro......... 125

2191 CAGACTGAAACCCCCTTGGAAGGCCTCAGTCAGGCAGCCCTAGAGACTG 2142
     ||||||||
 126 ..............GlySerPro........ 128

2141 GGGAGAGGAGAGGAGGACGACCCCCAGCTGTCAGCTACGCACCT 2092
     |||:::||    |||||||:::::::
 129 ........GlyAlaProGlyProGlyIleCysProGlyIleCysProGlnSerCysP 141

2091 CAGCAGCACAGGTGCCGAGCAGAGAGCCACATTAC.............. 2057
     ::|||||||    ::|||:::||
 141 roSerIleSerGlyGlyMetGlyTyrProGlnThrTyrAspSerTyrAspVal 157

2056 ...TTTGGCAACAGAAACTGGCGGCCAGCCCGGCAG............ 2021
     |||          :::|||||||
 158 LysAlaGlySerValGlyMetGlyProProGlnProProIleSerGlyPh 174

2020 .CCCCATGGGCGTAACAGAGGCGGGAGCTGGGA...CCCAGTGAGGCAG 1975
     :::          |||    |||::::::|||||
 174 eProGlyArgThrGlyAlaProGlyProGlyProGlyHisAlaG 191

1974 GCCCTCCACCCAAATGTCTGGAAGTTTCTACGCTGAGTATTTGGCCAA 1925
     ||||||||          |||||
 191 lyProPro.........GlySer.......... 195

1924 GTCGGCTCTTGTCAAATACTACCTGTGTAGCAAAGTAAATGGCGACCAGAC 1876
     ||||||          :::||||
 196 ..........AsnGlyTyrGlnGl 200

1875 ...CCAGGCCTGCGGCAGACACCATATAGGCAGTGACAGACTGGCTGAG 1829
     |||:::::::          ::::
```

```
              |||||
 200 yProProGlyGluProGlyGlnProGlyGlyProSerGlyProGlyProA 217

1828 CTGACAATGGAGCCCATAAACACAGGAGGATGGGCACCTGGGACAGCAGGA 1779
     |||||::||        ::::|||    |||||||||||    |||
 217 laGlyMetIleGlyPro.......AlaGlyProProGlyLysAspGly 230

1778 AGGCACTATCCAGGATGGCGAGGTGCCCGGCCCGGAAC 1730
     :::        |||||||    ||||||||||    |||
 231 Glu........ProGlyArg...ProGly.ArgAsnG 239

1729 ..........                CACCCTGGCCTC 1718
                               ||||||||||:::::
 239 lyAspArgGlyIleProGlyLeuProGlyHisLysGlyHisProGlyMet 255

1717 GGTGGGCTCACC...........                CACCACCA 1698
     |||:|
 256 ProGlyMetProGlyMetLysGlyAlaArgGlyPheAspGlyLysAspGl 272

1697 CACCTACGGAGCAATCACAGGCAGGAGGGCCCCGCAG........... 1663
     :::    ||||| |||    |||
 272 yAlaLysGlyAsp.SerGlyAlaProGlyProLysGlyGlyGluAlaGlyGln 288

1662 ..AGCGCCGGTGGAGTGGGAGCAGGCCACTGCTCCAGCACCCACCGTG 1616
     ::|||:::|||:::    |||    ::|||||||
 289 ProGlyAlaAsnGlySerProGlyGlnProGlyGlyProGlyProThrGl 305

1615 TCCATTAGGGAAG...GGAGCTCCAGGCTTAGGGCCT........GGCA 1578
     |||||||||          |||    ||:::
 305 yGluArgGlyArgProGlyAsnProGly.....GlyProGlyAlaHisGlyL 321

1577 GGAAGCTGGTCATCAGGCTGTCCTCACTGTGCTAGCACCTCCAGTGTCCCCT 1528
     :::        |||:::||::||||||
 321 ysAspSerGlyAlaProGlyThrAlaGlyProLeuGlyProGlyProPro 337

1527 CGGTATTTGGGCAGGAACACCTGCTTCTCCGGTGGTAGGGGAGG..... 1482
     |||        |||:::||||||    :::::||||
 338 .GlyThr...AlaGlyPheProGlySerProGlyGlyGlyGluAlaG 353

1481 ..........        CCAGTGTGTAGGGACGGATCTGCAGGTGCGTGAAGGTGA 1443
     :::        |||:::||    ||    |||    |||
 353 lyProProGlyProAlaGlyAlaSerGlyAsnProGlyGlyGluArg.GlyGl 369

1442 ACCCGGTGAGGGGCCTGAAGCTGTCACCGGCGCACACTGTGGGACAGG 1393
     |||||||          |||    :::|||
 369 uProGlyProGlnGly.....GlnAlaGlyProGlyProGlnProPr 385

1392 CATGTGGCACCGGCAGCCACAGGAAAGCTGCCACACTGGCCAAATAGAC 1343
     :::||    ||||||          ||:::||
 385 oGlyArgAlaGlyAlaGlySerProGlyLysGlyLysGlyGluMetGlyProSerGlyLyI 402

1342 TGCTCGAGTGCCGAATCCTCGCTGACCAGCCGCGGTCCATGACCA......... 1302
     |||        :::::||||||
 402 le..................ProGlyGlyGlyProGlyProProGlyGly 411

1301 ..........                GAGAGAAGAACCAGGAGGATGGCG 1279
     :::        ||:::||||||
 412 ArgGlyLeuProGlyProProGlyThrSerGlyAlaAsnProGlyAlaLysGlySerI 428

1278 CACTGCAGGAACAGCC...........                CCAGGCTGCC 1253
     |||        |||
 428 yThrProGlyGluProGlyLysLysAsnGlyAlaLysGlyAspProGlyProL 445

1252 CATCCGAACGCCTTCATCATAGTGTCTCCGGGCCTCCGGTGCCCGGCTCAG 1203
     ||||||||          |||||
 445 ysGlyLeuArgGlyGluAsnGlyTyrProGlyAlaArgGlyProProGly 461

1202 CCTCTGGGCACCGCCTGGTACAGCCCCTCGCCCCACGAAATCCGTGTAAAAC 1153
     ::::::        |||:::|||
 462 GluGluGlyLysArgGlyAlaAsnGlyGluProGly.............. 473

1152 AGGCTGAAGGTCATGAGTGCCATCCAGCTGCACAGCTCAGCACGAAGAG 1103
     :::::::|||||||:::::::    |||    :::
```

474 ......GlnAsnGlyValProGlyThr..........ProGlyGluA 484
1102 CCGGGCCA..............GGGTGCGGGGCATGCGGCAGC 1074
      |||::: ||| :::::||| :::
484 rgGlySerProGlyPheArgGlyLeuProGlySerAsnGlyLeuProGly 500
493 GAGCGGGACACAGACCAGGCCCAGCACTGGACCAATGCCCAGCACCATGG 444
      || |||||||||| ||||||||||
764 lyProIleGlyProProGlyProAla.....GlyProProGlyAspLys 778

443 TCATGAACTTCTCCTCTA...CCCCCACTTCCAGCAGCAGAGGCGGCACA 397
      ::: |||:||| ||| ||:::|
779 GlyGluThrGlyProAlaGlyAlaGlyProGlyProAlaGlySerArgGly.. 794
1073 ACAGCTGGTGCAGCCGGGAGCAGGCGCCCAGGTTCCGAAAGCCAAG 1024
      ||||||||||| ||| |||||||
501 GlulysGlyProAlaGlyLysGluArgGlyLeuProGlyProGlyProSe 517

396 TAGGTGATGCCTGCGGCCAAACACACCTCCAGGCCAAAGGTTAGCAGGTT.. 347
                              |||||:|||
795 ....GlyProGlyGluArg...........799
1023 CGGGCCCGGCCATGGACAGCAGTGGGCGGCACAAGAGGGGCCGACAGCCC 974
      ||||||::: ||| ::: ||| :::
517 rGlyProAlaGlyAsp..........ArgGlyGlnAspGly. 527

346 GACCAGCAAGAGCTGGGCTTTCCGGTGCCGCAGCAGGCGGCTCACCCACA 297
                   :::::| ||||| ||| |||
800 .....GlyGluGlnGlyLeuProGlyProAla.........GlyPheProGly 812
973 TTCTGCTGGCTCGGTGGGCCCCAGGCGTGCCT.............. 942
      ||||||||||| ||:||:||
528 ....GlyProGlyLeuProGlyMetArgGlyLeuPro 538

296 GCCTCTGACCATAGTGGG...CCAGGCGGGTAGGGTTCAGGGGCCCGTT 250
      ||| |||||||||| ||||||
813 AlaProGly.GlnAsnGlyGluProGlyGlyLysGlyGluArgGlyProP 829
941 ..............CCTCAGGCCACCAGCA.. 927
                            ||| ||||||
539 GlyIleProGlySerProGlySerAspGlyLysProGlyProProGlyAs 555

.GTGTGGCTGCTACGCAGGTGAGGAAGATGAGGGTGAGCAGCCAAAGAG 878
926 ||||||
249 CAGGCACTCCAGAACTGCTTCGTCCGCTCGTCCTGCTCCAGAAGCTGCGCC 200
      ||| ||| ||||| :::||||||
555 nGlnGlyGluProGlyArgSerGlyProProGlyProAlaGlyProArgG 572
829 roGly..............LeuArgGlyLeuAlaGlyProProGlyAlaAla 841

877 GCACTCCT.....................CCTGGGTGC 861
      ||::: |||::||
572 lyGlnProGlyValMetGlyPheProGlyProLysGlyAsnGluGlyAla 588
199 TCTCCCTCCTTGCTGCCGCCAACTGCCTAGGAATCAGCCAGGCGCCCATTT 150
      ::: ||||||||
842 Gly..............................ProGlnGlyGlyProGlyAla..... 849

860 CCAGGT.....AGGGGGCCAGGCACTGGTCCCATGTCCAGTCAATG 823
      ||||||| ||||||||| :::
588 ....AGGGGGCCAGGCACTGGTGTCCCAGTCAATG 823
149 CTGCCAGCCCTTTGGTGCCGGTCCAGCTTCTCAGCCCATGCTCAACACCT 100
                        ||||||
850 ....ProGlyProPro.................. 853

589 ProGlyLysAsnGlyGluArgGlyLeuProGlyGlyLysProGlyThrProGl 605
822 GCAGGCAGGA............GGTAGCCCAGGCAGCCCCCAAGACT 788
      ||||| ||| |||
99 GCTGCTGTGGGGCACCTCAGTGGGACACGTCTCATCACTCAGATCCT.. 52
854 ....GlyProGln.GlyValLysGlyGluArgGlySerProG1 866

605 yProAlaGlyLysAsnGlyAspValGlyLeuLeuGlyProProGlyProA 622
787 GATCATGAAGGCATAGACAGAGTAGGCCTGCCACGT............. 750
      ::::::||||||:|||::|||||||
51 .GCCCGAGGCGCGCGGCGTGCTCACCCGGAGCC 22
866 yGlyProGlyAlaAlaAlaGlyPheProGlyGlyAla 876

622 laGlyProAlaGlyAspArgGlyGluProGlyProSerGlySerProGly 638
749 ..............GTCCGGGTCCCCGAAGAGGTCAGAGAGCA.. 720

749 ..............GTCCGGGTCCCCGAAGAGGTCAGAGAGCA.. 720
       :::::::::|||||| ||:::::::
639 LeuGlnLeuProGlyLeuProGlyProGlyProAla...GlyGluAsnGlyLy 654

719 ....G 719
                         |
719 ....ProG 716

654 sProGlyGluProGlyLysGlyLysGlyAspIleGlyGlyProGlyPheProG 671
718 GGGCTCCAGTGGAGTGAAGCACACCTGGCCACAGAAGTCCAGCAGCCCCA 669
      |||| :::::||:||: ::|||
671 lyProLysGlyGluAsnGlyLylleProGlyGluArgGlyProGlnGlyPro 687

688 ProGlyProThrGlyAlaAlaArgGlyGlyProGly........ProAlaG1 701
668 .......CGCCCAGGATGAGCAGTGCCAGCT.. 645
                ||||||
668 ........CCAGGGGCCGCTGGGATCCGG 626

644 yProThrGlyAlaAlaLysGlyProGlyProGlyAla.....ProG 716
       ||||||||| ||| |||
701 ySerGluGlyProGlyProLysGlyAspIleGlyLysGlyProProGlyPheProG 671

625 GCACAGCAGCCCTGCTAGCCAGCCGCCCTTGGCATGAGAGAGAGAGGCTCA 576
      || :::::|| ::: |||||||||||:::
716 lyGlyThrGlyLeuProGlyLeuGlnLysGlyMetProGlyGluArgGlyAla 732

575 GCAGGATGCCCAAGGACAGTGCCACGATGAGAGGGCCGGC.....GGCGG 532
      ||||||||||| ::: ||||||
733 SerGlySerProGly.....ProLysGlyAspLysProLysGlyGluProGlyGl 747

531 CCATAGCGTCCAC............GCCAGTGGTCACTGGCCTAGCCTAG 494
      | ||| :::|||
747 yLysGlyAlaAspProLeuProGlyAlaArgGlyGluArgGlyAsnValG 764

seq_name: sp_rodent:Q60444

seq_documentation_block:
ID   Q60444   PRELIMINARY;   PRT;   1549 AA.
AC   Q60444;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   TYPE VII COLLAGEN (FRAGMENT).
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE; 93271985.
RA   GREENSPAN D.S.;
RT   "The carboxyl-terminal half of type VII collagen, including the
RT   non-collagenous NC-2 domain and intron/exon organization of the
RT   corresponding region of the COL7A1 gene.";
RL   Hum. Mol. Genet. 2:273-278(1993).
DR   EMBL; L06863; AAA36968.1; -.
DR   PFAM; PF01391; Collagen; 19.
DR   PFAM; PF00014; Kunitz_BPTI; 1.
DR   PROSITE; PS00280; BPTI_KUNITZ; 1.
KW   Structural protein; Serine protease inhibitor.
FT   NON_TER   1   1
SQ   SEQUENCE   1549 AA; 148117 MW; 3B72E093 CRC32;

alignment_scores:
      Quality: 335.50   Length: 1245

us-09-030-606-110.rspt

Wed Sep 29 14:26:57 1999

```
                    Ratio:    0.676              Gaps:        81
Percent Similarity:  39.839    Percent Identity:  26.024

alignment_block:
US-09-030-606-110/rev x Q60444  ..

Align seg 1/1  to: Q60444  from: 1  to: 1549
```

```
3039 GGAGCTGGTGGGGAAAGTTGGGGTAGGGAAGTTGGGGTAGGGAAA 2990
     |||||::||| |||  :::         :::|||:::
  36 GlySerProGlyProGlnGlyLeuProAlaGlyArgAlaGlyGluLysGlyGl 52

2989 TTTTGGCAGTGCCTTCAT................CAGCCC.... 2965
     : |||:|||                         ||| |||
  52 uLysGlyAspCysGluAspGlyAlaProGlyLeuProGlyGlnProGlyA 69

2964 ............................AGTCCTAGA 2956
                                 :::||||||
  69 laProGlyGluProGlyLeuArgGlyThrProGlyIleThrGlyProLys 85

2955 GAGAGTAGAGGGGAGTGGAAGTGGGGGAACCAGGCTGGCCAAGAGAAG 2906
     ::|||||||  |||            |||  |||||| ||
  86 GlyAspArgGlyGln.....ThrGlyThrProGly.GluProGlyGluL 100

2905 AGGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCCCCACCCTTCACCTTC 2856
     :::|||  ||| |||  ||::: |||:::
 100 ysGlyGluArgGlyProPro.GlyProValGlyProGlnGlyLeuProGl 116

2855 CTTCAACACCCTAACCTTGGGTAACAGCATTTGAATTATCATTTGGGAT 2806
     :::    |||         |||  |||
 116 y.......ValAlaGly.........HisProGly. 122

2805 GAGTAGAATTCCAAGGTCCTGGGTTAGGCATTTTGGGGGGCCAGACCCC 2756
     |||||||:::                   |||  ||| |||
 123 .ValGluGlyProGlu.............GlyProGlyProPro 132

2755 AGGAGAAGAAGATTCTGGCAATGACCAATCAGCTATCTCAGG 2706
     ||||||||   ::  |||
 133 AlaGlyArgArgGlyLeuValGlyGluProGlyArgPro......... 145

2705 GGACCTGATTGTTGGGGATCCCCACCCTACCCA...AATATTAGACACC 2659
     |||||||| ||: |||  :::
 146 .........GlyAspProAlaValGlyProGlyGlyAlaGlyAlaL 158

2658 AACACAGAAAAGCTAGCAATGGATTCCCTTCTACTTTGTTAAATAAATAA 2609
     ::|||:||||| |||::::|||
 158 ysGlyGluLysGlyAspAlaGlyLeuPro................. 167

2608 GTTAAATATTTAAATGCCTGTGTCTCTGTGATGGCAACAAGGACCCAAC 2559
     |||||
 168 .........................GlyProAr 170

2558 AGCCACATCCTGATAAAAGGTAAGAGGGGGTGGATCAGCAAAAAGACA 2509
     |||         |||||||:::|||
 170 gGlyAlaAlaGlyIleLeuLysGlyGluGlnGly......... 180

2508 GTGCTGTGGGCTGAGGGACCTGGTTCTTGTGTGTTGCCCCTCAGGACTC 2459
     |||  ||||||  :::
 181 ..................ProProGlyLeu 184

2458 TTCCCCTACAAATAAGTCATATGTTCAAATCCCATGAGGAGAGTGTTTCAT 2409
     ::||:|  :::::||| ::
 185 .......AlaLeuProGlyAspAlaGlyLeuLys......... 193

2408 CCTAGAAACTCCCATGCAAGAGCTACATTAAACGAAGCTGCAGGTTAAGG 2359
     |||||||:::||||::::|||
 193 .........GlyAspProGlyLys.......... 193

2358 GGCTTAGAGATGGGAAACCAGGTGACTGAGTTTTATTCAGCTCCCAAAAAC 2309
     |||     |||||||||
 194 ..........GlyAspProGlyAsp............ArgGly 200

2308 CCTTCTCTAGGTGTGTCAACTAGGAGGCTAGCTGTTGTTAACCCTGAGCCT 2259
     :::|||
```

```
 201 Pro...IleGlyLeuThrGlyArgAlaGly.........ProThrGlyAs 213

2258 GGGTAATCCACCTGCAGAGTCCCGCATTCCAGTCGATGGAGCCCTTCTG 2209
     ::::|||||:::||||::::||:           :::|||  ::|||
 213 pSerGlyProProGlyGluLysGlyAspProGlyArgProGlyProProG 230

2208 GCCTCCCCTGTATAAGTCCAG.................ACTGAAACCCCC 2177
     || |||||||                          ::
 230 ly...ProValGlySerArgGlyArgAspGlyGluValGlyGluLysGly 245

2176 TTGGAAGGCCTC.......CAGTCAGGCAGCCCTAGAGACTGGGAGAGAG 2133
     :::|||||                ||   |||||||||:::|||||
 246 ValGluGlyAsnProGlyLeuProGlyLysAlaGlyGluArG 262

2132 G......AGAGGACGCGCC....CAGCCCCCAGCTGTGCAGCTACGCA 2095
     |       ||||||          ||||||||||||
 262 gGlyLeuArgGlyAlaProGlyValArgGlyProAla.......... 274

2094 CCTCAGCAGCACAGGGTGCAGCAGAGACCACCACATTACTTTGGCAGCAAC 2045
     ||| :::|||  :::|||         ||  |||
 275 .....GlyGluLysGlyAspGlnGlyAspProGlyGluAspProGlyArgAsn 289

2044 AGAAACTGGCGGCCAGCCCGGCAGCCCCATGGGGCTAACAGGAGCGGGGA 1995
     |||||||||||||||||||            ::: |||     |||||
 290 GlySerProGlyProSer...GlyProLysGlyAspArg......GlyGl 303

1994 GCTGGGACCCAGTGAGGCAGGCCCTCCACCCCAATGTGCTGGAAGTTTC 1945
     :::|||||   |||||||:::||||                |||
 303 uProGlyPro.....ProGlyProGlyArgLeuValGlyAlaGlyGlyI 318

1944 TACGCTGAGTATTTGGCCAAGTGCCTGCTCTTGTCAAATACTACCTGTGTAGC 1895
                                            :::
 318 leGly.........................................Ser 320

1894 AAAGTAAATGGCGACCA.............................GA 1876
     :::::::                                       ||
 321 ArgAspLysGlyGluProGlyGlnGluGlugluProGlyProGlyLysGlyAs 337

1875 CCCA........................GGCTGCGGC 1862
     |||                         |||||||||
 337 pProGlyProProGlyAlaSerGlyGluArgGlyIleGluGluLeuArg 354

1861 AGACACCATATAGGCAGTGACAGACTGGCTGGACAATGGAGCCCA 1812
     |||      |||  |||::   :::
 354 lyProProGlyProGlnGlyAspProGlyValArgGlyProAlaGlyAsp 370

1811 TAAACAGGGATGGGGCCACCTGGGACAGCAGGAGGCACTATCCAGGATG 1762
     |||||||||||||||:::          ::
 371 LysGlyAspArgArgGlyProProGlyLeuAspGlyArgAsn......GlyVa 385

1761 GCGAGGTCCAGCAGATGCCCCGCCCGGAACCACCCTGGCCTCGGTGG 1712
     |||         ::|||||||||||
 385 lAspGlyLysProGlyAlaProGlyPro.................... 394

1711 CTCACCCACCACCACTACGGAGACATCACAGGACAGGAGCCCCGCAGA 1662
     :::||||||||  :::::||: :::|||
 395 .....ProGlyProHis..GlyAlaSerGlyLysAlaGlyAspProGlyA 409

1661 GCGCGGGTGGAGGTGGGAGCAGG.........CCACTG...CCTCCA 1627
     |||      |||  |||||
 409 rgAspGlyLeuProGlyLeuArgGlyGluHisGlyGlyProProGly 425

1626 GCACCCAGGTGTCCAGGCTTCATTAGGGAAG...GGAGCTCCAGGGTT 1589
     ||| |||||||               |||      ||||||||
 426 GlyProProGlyValProGlyLysThrGlyGluAspGlyLysProGlyLe 442

1588 AGGG.......................CCTG 1581
     ::::                       ||
 442 uAsnGlyLysAsnGlyGluProGlyGluAspProGlyLyArgGlyLysG 459

1580 GCAGGAAGCTGGTCATCAGGCTGTCCTAGCACCCTCCAGTGTCC 1531
     |||:::|||||||||:::||||
```

us-09-030-606-110.rspt

```
459  lyGluLys.GlyAspSerGlyAlaProGlyArgGluGlyProAspGlyPr   475
1530 CCTCGGTATTGGCAGGAACACCTGCTTCTCCCGTGTAGAGGGAGGC        1481
475  oLysGly..GluArgGlyAlaProGlyAspProGlyLeuArgGlyProP    491

1480 CA........GTGTGTAGGGCAGGATCTGCAGGGCTGAGAAGGTGA        1443
491  roGlyLeuProGlyGln.ValGlyProProGlyGlnGlyPheProGlyVa   507

1442 ACCCGGTGAGGGCCGGTGAAGCTGTCACCACGGCCACACTGTGGACA..     1395
507  lProGlyAsnThrGlyProLysGlyAspPArgGlyGluThr..GlySerLy  523

1394 ................GGCATGTGGCACCGGCAGCCAC                1373
523  sGlyGluGlnGlyLeuProGlyGluArgGlyLeu.....ArgGlyGluP    538

1372 AGGGAAAGCTGCCACACTGGCCAAATAGACTGCTCGAG......TGC       1332
538  roGlySerLeuProAsn...AlaGluArgLeuLeuGluThrAlaGlyIle   553

1331 CGAATCGCTGCA..........................CCAGCCGG       1312
554  LysValSerAlaLeuArgAspIleValGluThrTrpGlyluSerSerGl    570

1311 T..............CCATGACCAGAGAAGACCAGGGAGATGGCGCACTGCAG 1271
570  ySerPheLeuProValProGluArgProGlyProLysGlyAspProG      587

1270 GAACA..........................GCCCCAGGCTGCCCATCCGAA  1245
587  lyGluArgGlyProProGlyLysGlyGluGlyLySerIleGlyPheProGlyGlu 603

1244 CGCCTTCATCATAGTGTCTCCGGGCCTCGGTGCCCGGCTCCAGCTCTGGGC   1195
604  Arg.......GlyLeuLysGlyAspArgGlyAspProGlyArgGlyGlnGl   617

1194 ACGGCCCTGGTACAGCCCTCGCCCACGAAATCCGTGTAAAACAGCGTGAA   1145
617  yProProGlyLeuAlaLeuGlyGluGlyProProGly.............   630

1144 GGTCATGAGTGCCATCCAGCTGCACAGCGTCACACGCCACGAAGAGCCCGCA 1095
631  ............ProProGlyLeuAlaGlyGluProGlyLyaProGlyIle  643

1094 GGGTGCGGGGCATGCGGCAGCAGCTGGTGCAGCCGGGAAGCAGGGCG       1045
644  Pro...GlyLeuProGly...ArgAlaGlySerAlaGlyGluAlaGlyAr   658

1044 CCCAGGTTCCGGAAAGCCAAGCGGCCCCGCCATGACAGCAGCAGTGGGGCGA  995
658  gProGlyGluArgGlyGlyLysGluLysGlyAspArgGlyGluGlnG      675

994  CAAGGAGGGGGCCGCCGACAGCCCTTCTGCTGCTCGGTGGGGCCCAGCCTG   945
675  lyArgAspGlyArgGlyLeuProGly..........GlyProProGly     687

944  CCTCCTCAGCCACCAGCAGTG.........TGGCTGCTACG            913
688  ProProGlyProLysValAlaIleAlaAspGluGlnGlyProGlyLeuSerA  704

912  CAGGTGAGGAAGATGAGGTGAGCCGGTGCAGCAGGCCACTCCT........  870
704  gGluGlnGlyProProGlyLeuLysGlyAlaLysGlyGluProGlySerA   721

869  ...............CCTGGTGCCCAGGTAGGGGGCCA              846
721  spGlyAspHisGlyProLysGlyAspLysGlyGlyAlaProGlyIleLysGly 737

845  GGGGCACTGTGTCCCAGTCAATGCAGGCAGGAGGTAGCCCAGGCAGCCC     796
737  lyArgAspGlyArgGlyLeuPro.........GlyProProGlyLeuPr    754

738  AspGlnGlyGluProGlyLysArgGlyHisAspGlyLySerProGlyLeuPr  754
```

```
795  CCAAGACTGATCATGAGGCATAGACAGAGTAGGCCTGGCGACAGTGGT      747
754  oGlyGluArgGlyValAlaAlaGlyProGluGlyLysProGlyLeuGlnGlyP 771

746  ........CCGGGTCCCGGAAGAGGTCAG....AGAGCAGGGCCTCCA      711
771  roArgGlyThrProGlyProAlaGlyHisGlyAspProGlyProPro       787

710  GTGGAGTGAAGCAGCACACCTGGCCACAGAAGTCCAGCAGCCCCACGC..   666
788  .........GlyAlaProGlyLeuAlaGlyProAlaGlyGlnGlyPr      801

665  .....................................CCAGGATGAGCAG   653
801  oSerGlyLeuLysGlyGluProGlyGluThrGlyGluProProGlyArgGlyL 818

652  TGCCAGCTCCA...GGGGCC...................TGGGATCCGGGCACAGCA 618
818  euProGlyProThrGlyAlaValGlyProGlyProGlyProSer          834

617  GCCCTGCTAGCCAGCCGGCCCCTTGGGATGAGAAAGAGGCTCAGCAGGATG   568
835  GlyLeuValGly.........................ProGlnGlySerProGlyLe 845

567  CCCAAGGACAGTG.................................CCCAGATGAAGGGCCG 539
845  uProGlyGlnValGlyGluThrGlyLyLysProGlyProProGlyArgAspG  862

538  GCGGCGCGCCATAGCCTCCACGCCAGTGGTCACTGGCTGCTGAGCCTAGGAGCG 489
862  ly...................ThrSerGlyLys........................Asp 867

488  GGACACAGACCAGGCCGCACTGGACCAATGCCCAGCCACCACCATGGTCATG 439
868  GlyGluArgGlyGlyGlyProGlyValProGlyLeuProGly            880

438  AACTTCTCCTCTACCCCCACTTCCAGCAGCAGAGGCGGCCACATAGGTGAT   389
881  ...LeuProGlyProValGlyProLysGlyGluProGly...............    892

388  GCCTCGCGGCCAAACACACCTCCAGGCCCAAAGTTAGCAGGTTGACCAGCA   339
893  .........ProValGlyAlaProGlyGlnValMetValGlyProProGly   906

338  AGAGCTGGGCTTTCCGGTGCC...............GCAGCAGGC...GG    307
907  AlaLysGlyGluLysGlyAlaProGlyAspLeuAlaGlyAspLeuLeuGl    923

306  CTCACCCACAGCCTCTGGACCCATAGTGGG.........................    278
923  yGluProGlyValAlaLysGly.AspArgGlyLeuProGlyArgGlyGlu    939

277  ....................................CCAGGCGGCGGGTAGGGCTCAGGGGG 255
940  LysGlyGluAlaGlyArgAlaGlyGluProGlyAspProGlyGluAspGl   956

254  CCGTTCAGGCACTCCAGAACTGCTTCGTCTCCGGC                  221
956  yGlnLysGlyGlyAlaProGly.....ArgLysGlyLeuLysGlyGluProG 971

220  .............TCTGCTCCAAAGCTGCGGCCTCTCCTCCTTGCTCTCCGC  182
971  lyIleGluValGlnGlyProProGlyProThrGlyProProGlyMetLys   987

182  ........................................              182
988  GlyAspValGlySerProGlyAlaProGlyValValGlyPheProGlyGl  1004

181  .CAACTGCCTAGG...............AATCAGCCAGGCGCCCCATTTCTGCCCAGC 142
1004 nThrGlyProArgGlyGluThrGlyGlnProGly.ProValGlyGluArg  1020
```

```
510 spGlyLeuProGlyProProGlyThrMetLeuPro.........         523
    ::|:::|
2834 TAACAGCATTGGAATTATCATTTGGGATGAGTAGTAGAATTCCAAGGTCCT  2785
    ::: |||||
524 .........PheArgPheGly.............................    527
     |||||
2784 GGGTTAGGCCATTTTGGGGGCCAGA..........CCCCCAGGAGA        2750
    || :::                          ||||||
528 .GlyGlyGlyAspAlaGlySerLysGlyProMetValSerAlaGlnGlus    544
    || ::::::||||||||||                   ||||
2749 AGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGA...    2703
    ::::: ||||||||||||
544 erGlnAlaGlnAlaIleLeuGlnGlnAlaArgLeuAlaLeuArgGlyPro    560
                                    CCTGATTGTTGGGGATCCCCCAC  2680
2702 .................................                     |||
561 AlaGlyProMetGlyLeuThrGlyLeuArgProGlyProMetGly.ProProG  577
                                       ::|
2679 CCTACCCAAATATTAGACACCAACACAGAAAAGCTAGCAATGGAATTCCCT   2630
    :::::::::||          ::|
577 lySerGlyGlyLeuLysGlyGluProGlyLeuProGlyGlnGlyPro        593
                      ||:          ||||||
2629 TCTACTTTGTTAAATAAATAAGTAAATATTAAATGCCTGTGTCTCTGT      2580
                                             :::||
594 Arg.................................GlyVa            596
2579 GATGGCAACAGAGGACCAACAGGCCACATCCTGATAAAAGGTAAGAGGG     2530
    |:::::  |||||||||||        ::::::::||
596 lGlnGlyProProGlyProThrGlyLysProGlyArgArgGlyArgGAlaG   613
2529 GGGTGGATCAGCAAAAGACAGTGCTGTGGGCTGAGGGGACCTGGTTCTT    2480
    || |||||||
613 lySerAsp..........................................    615
2479 GTGTGTTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAA   2430
615 ..................................................    615
2429 TCCCATGGAGGAGTGTTTCATCCTGAGAAACTCCATGCAAGAGCTACATT   2380
615 ..................................................    615
2379 AAACGAAGCTGCAGGTTAAGGGCTTAGAGATGGGAAACCAGGTGACTGA    2330
                 ||| ||| |||||||
616 ...............GlyAlaArg...GlyMetProGlyGln...         623
2329 GTTTATTCAGCTCCCAAAACCCTTCTCTAGGTGTGTCAACTAGGAGG      2280
    :::::::||||||||    :::  ||| :|||||:||||||||
624 ......ThrGlyProLysGly..AspArgGlyPheAspGlyLeuAlaGl     637
2279 CTAGCTGTTAACCCTGAGCCTGGTAATCCACCTGCAGAGTCCCCGCATT    2230
    |                |||::::||| ||||||||||
637 y.............................LeuProGlyGluLysGlyHisA   645
2229 CCAGTGCATGGAGCCCTTCTGCCTCCGTATAAGTCCAGACTGAAACC      2180
    :::                      ::|||||||||
645 rgGlyAspProGlyProSerGlyProPro.....................    654
2179 CCCTTGGAAGGCTCCAGTCAGGACGCAGCCTAGAGACTGGGGAGGAGG     2132
    ::::|||||                 ||||||||||
655 ...............GlyIleProGlyAspAspGlyGluArgGl          664
2131 ...............AGAGGGACGCGCCCCCAGCCCCCAGCTGTGC        2104
                    ||||||      |||:  ||||||
664 yAspAspGlyGluValGlyProArgGlyLeuProGlyGluProGlyProA    681
2103 AG....CTACGACCTCAGCCAGCCAGGTGGCAGCAGAGCCACAT         2060
    ::                ::::|:        |||
681 rgGlyLeuLeuGlyProLysGlyProProGlyProProPro...          696
2059 TACTTTGGCAGCAACAGAAACTGGCGGCCAGCCCGGCCCCATGGGGC      2010
    |||:::|||                    ||||||||  |||   |||
697 ......Gly.ValThrGlyMetAspGlyGlnProGly...ProLysGlyA    710
```

```
141 CCTTTGGTG.................................CCGGTCCAGCT    122
    |||:||
1021 GlyLeuAlaGlyProProGlyProGlyArgGluGlyAlaProGlyPro.....  1034
121 TCTCAGCCCATGCTCAACACCCTGCTGTGGGG......CACCTCAGTG      78
    |||::|||    :::::||::::||||||                  |||||
1035 .LeuGlyProProGlyProProGlySerValGlyValGlyAlaProGlyAlaSerG  1051
77 GGGACACGTCTCATCACTCAGATCCTGGC.................CGA    46
    ||   :::  :::||||||||||                   |||
1051 lyLeuLysLysGlyAspLysGlyAspProGlyThrGlyLeuProGlyProArg  1067
45 GGCGCGGCGGCTGTCACCCGGA                               25
    |||||:|:|||||||||
1068 GlyGluArgGlyGluProGly   1074
```

```
seq_name: sp_rodent:O88207

seq_documentation_block:
ID   O88207    PRELIMINARY;    PRT;  1838 AA.
AC   O88207;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   COLLAGEN A1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE: 98250615.
RA   WU Y.L., SUMIYOSHI H., KHALEDUZZAMAN M., NINOMIYA Y., YOSHIOKA H.;
RT   "cDNA sequence and expression of the mouse alpha1(V) collagen gene
RT   (Col5a1).";
RL   Biochim. Biophys. Acta 1397:275-284(1998).
DR   EMBL; AB009993; BAA28786.1; -.
DR   PFAM; PF01410; COLFI; 1.
DR   PFAM; PF01391; Collagen; 17.
SQ   SEQUENCE  1838 AA;  183691 MW;  05114C5E CRC32;

alignment_scores:
                Quality:   335.50        Length:    1213
                  Ratio:     0.692          Gaps:      78
Percent Similarity:   39.984      Percent Identity:   24.815

alignment_block:
US-09-030-606-110/rev x O88207

Align seg 1/1  to: O88207  from: 1 to: 1838

3068 GGTCCTGCAGTAGCTCCAAACAGGGTTGTGGAGCTGGTGGGGAAAGTTGG   3019
    ||||||::||:::|||      ::: :|::||||         |||||
431 GlyProGlyMetProAlaAsnGlnAspThrIlePheGluGlyYIleGlyGl   447
3018 GGGTAGGGGAAAGTTGGGG...GTAGGGGAAATTTTCGGCAGTGCCTTCA   2972
    ||   ||||||   |||||      :||||:||| |||:::|          |||:::||
447 yProArgGlyGluLysGlyGlnLysGlyGlu........ProAlaIleI    461
2971 TCCAGTCCAGTCCTAGAGAGAGTAGAGGGG......................  2943
    |||::::||| :::                      ||||||||||
461 leGluProGlyMet...LeuIleGluGluProGlyThrThrGlyLysGlyPro  476
2942 .................................AGTGGAAGTGGGGGGGGAACCAGGC  2920
                                   ::||||  ||:||||||||
477 AlaGlyLeuProGlyProProGlyThrThrGlyProThrGlyGlnMetGl   493
2919 TGGGGCCAAGAGAAGAGGGGGTGGTTAGGGGAAGCCGGTTGAGACCTGAAGCC.   2871
    ||| ::::::|||||::|||      |||   |||
493 yAspProGlyGluArgGlyArgProGlyArgProGlyLeuProGlyAlaA   510
2870 .................CCACCCCTTACCTTCCTTCAACACCCTAACCTTGGG   2835
    ||||||::|||||:::|||:::: ::|||:: |||
697 ......Gly.ValThrGlyMetAspGlyGlnProGly...ProLysGlyA   710
```

```
2009 TAACAGGA..GCGGGGAGCTGGACCAGTGAGGCAGGCCCTCCACCCC 1963     1274 GCAGGAACAGCCCA...GGCTGCCATCCGAACGCCTTCATCATAGTGT 1228
          ::|||    |||::: ||||||||||              ::||||              :::||| ::::||||  ||| |||:::||:::|||||
 710 snValGlyGluProGlnGlyProGlnGlnGlyAsnPro 726               937 GlyGlyAspGlyProAlaGlyProGlyGluArg.......... 948

1962 AATGTGCTGGAAGTTTCTACGCTGAGTATTTGCCAAGTCGCTCTTGTC 1913     1227 CTCCGGGCCTCGGTGCCGGCTCAGCTCTGGGCACGCCCTTGGTACAGCC 1178
          :::        |||        ||| |||      :::||||||               |||||||:  |||      :::  |||||||    ||||||
 727 Gly.........AlaGlnGlyLeu.ProGlyProGlnGlyA 737             949 ....GlyProAsnGlyProGlnGlyProThrGlyPheProGlyProLysG 964

1912 AAATACTACCTGTGTAGCAAAGTAAATGGCAACAGACCCAGGCCTGCGG 1863     1177 CTCGCCCACGAAATCCGTGTAAAACAGCGTGAAGGTCATGAGTGCCATCC 1128
          ||| |||  ::::::              ||||||||||                    ||||||||||              :::
 737 IaIleGlyProProGlyGluLysGlyProLeuGlyLysProGlyLeuPro 753      964 lyProProGlyProProGlyLys.............Asp 972

1862 CAGACACCATATAGGCAGTGACAGACTGGCTAGCTGGACAATGGAGCCC 1813     1127 AGCTGCACAGCTCAGCCACGAAGAGAGCCGGCGCAGGGTGCGGGCATGCGG 1078
          |||   :::  ||  ::|||                                      :::             |||  |||||
 754 GlyMetProGlyAlaAspGlyProProGlyHisProGlyLys........ 767      973 GlyLeuProGlyHisPro............GlyGlnArgGl 982

1812 ATAAACAGGGATGGGGCCACCTGGG...ACAGCAGGAAGGCACTATCCAG 1766     1077 CAGCACAGCT.....GGTGCAGCCCGGGGAAGCAGGGCGCCCAGGTTCCG 1034
          ||||||||||           :::||||   ||||                       |::||||::    ||| ::||| ||| |||||
 768 ........GluGlyProProGlyGluLysGlyGlyGlnGlyProProG 781       982 yGluThrGlyPheGlnGlyLysThrGlyProProGlyProGlyValV 999

1765 GATGGCGAGGTCCAGCAGCAGATGCCCCGGAA.......... 1731          1033 GAAAGCCAACGCGGGCCCGGATGGACAGCAGTGGGGCAGAAGGAGGGGG 984
          ||  :::|||| |||||||||                                    ||:::||||||              ::: |||
 781 lyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 797      999 alGlyProGlnGlyPro............ThrGlyGluThr......Gly 1009

1731 .......... 1731                                          983 CCGACAGCCCTTCCTGGCTCGGTCGGTGGGCCCAGCGTGCCTCCTCCAGCC 934
                                                                       |||            :::::||| |||  ::|||   ||||:::
 798 AspGlyIleArgGlyLeuLysGlyThrLysGlyGluLysGlyGluAspGl 814      1010 Pro..........MetGlyGluArgGlyHisProGlyProGlyPr 1022

1731 .......... 1731                                          933 ACCAGCAGTGTGGCTGCTCTACCCAGGTGAGGAAGATGAGGGTGAGCAGGCC 884
                                                                       |||||::         :::  |||  |||||:  ||||||
 814 yPheProGlyPheLysGlyAspMetGlyIleLeuGlyLysGlyAspArgGlyGluI 831  1022 oProGly............GluGlnGlyLeuProGlyAlaAlaAlaGlyL 1034

1730 ...CCACCCTGGCCTCGGTGGGCCTCACCCCACCACCACGTACGGAGA 1686    883 AAAGAGGCACT.....CCTCCTGGGTGCCCAGGTAGG.... 852
          ||||||| ||||| |||:::                                      :::||||:::        |||   ||||||
 831 leGlyProProGlyProArg..GlyGlu............ 839            1034 ysGluGlyThrLysGlyAspProGlyProAlaGlyLysLeuProGlyLysLysAsp 1050

1685 CATCACAGGCAGAGGCCCCCAGAGACGCGGGTGAGGTGGGAGCAGGCCA 1636    851 GGGCCA.......... 826
          |||::::||:::||  |||||   :::|||                            |||||||
 840 ..AspGlyProGlyLeuProLysGlyLyArgGlyGlyProAsnGlyAspPro 855     851 GlyProProGlyLeuArgGlyPheProGlyAspArgGlyLeuProGlyPr 1067

1635 CTGCCTCCAGCACCACAGTGTCCATTAGGGAAG..GGAGCTCCAGGCTT 1589   825 ATGGCAGGCA.......... 794
          ||| ||  :::|||||||||         :::||::||||||||                |||
 856 GlyProLeuGlyProThrGlyLysGluLysGlyLysLeuGlyValProGlyLe 872     825 GlyGluSerGlyLeuLysGlySerGluGlyProGlyProProG 1084

1588 AGGG.....CCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAG 1545   793 AAGACTGATCATGAAGGCATAGACAGAGTAGGCCTGGCACAGTGTCCG 744
          |::                ||||||||||||                         :::          ||||||:::
 872 uProGlyTyrProGlyArgGln................G 880                793 lyPro..........AlaGlySerProGlyGluArgGlyPro 1094

1544 CACCTCCAGTGTCCCCTCGGTATT.....TGGGCAGGAACACCTGCTTC 1501   743 GGT.......... 718
          ::||| ||| ::|||:::       :::  ||||||                        |||||||
 880 lyProLysGlySer.IleGlyPheProGlyAlaAsnGlyGl 896              743 AlaGlySerAlaGlyProIleGlyIleProGlyArgProGlyProGlnGl 1111

1500 TCCCGGTGGTGAGAGGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGA 1451  717 GCCTCCA..........GTGGAGTGAAGCACACCTGCCAAGGACAGTCCAG 677
          :     |||||||||||:::|||||:::|||                           |||||||         ||||||    :::||||||
 896 uLysGlyGlyArgGlyThrPro..........GlyLysProGly.... 907         717 yProProGlyProAlaGlyGluGluLysGlyGlnLysGlyGlyProG 1128

1450 GAAGGTGAACCCGGTGAGGCGGCGTGAAGCTGTCACCACCGCCACACTGT 1401  676 CAGCCCCACCGCCAGGATGAGCAGTGCCAGCTCCAGGGGCCTGGGATCCG 627
          |||           :::||||||   :::    ||||||||||               ::||| |||        :::|||||||||
 907 .........ArgGlyProThrGlyProArgGlyG 912                        676 lnGlyProLeuGlyProGly.Pro...GlnGlyProIleGlyProG 1141

1400 GGGACAGGCATGTGGGCACCCGCAGCCAGCAGGGAAAGTCGCCACACTGGCC 1351  626 GGGCAGCAGCCTGCTAGCCACCGCCCTTGGGATGAG.......... 587
          ::|||:::||:::        |||:::||:::                            :::|| ||| |||::::||||||
 908 ........ProArgGlyGln.......... 911                         626 GlyGluAlaGlyValGlyProGlyGlnGlyArgAsp.......... 1094  (??)

1350 AAATAGACTGCTCGAGTGCCGAATCGCTGCCACCACGTCCATGACCAG 1301  586 ..........AAAGAGGCTCAGCAGGATGCCCAAGGACACAGTGCCCA 551
          :::|||:::|||                                                ::||||:::                 |||:::
 912 ..........ArgGlyProThrGlyProArgGlylyG 920              586 yLeuProGlyProAlaGlyProValGlyProProGlyGluAspGlyAspL 1158

1300 AGAGA..........AGACCAGGG...AGATGGCCACT 1275       551 ..........GTGGAGTGAAGCACACCTGCCAAGGACAGTCCAG 551
          |||||                  :::  ::|||||||||                  :::|||||      ||| :::  |||:::|||:::|||
 920 luArgGlyProArgGlyIleThrGlyLysProGlyProLysGlyProGlyLysGlyAsnSer 936  1158 ysGlyGluIleGlyGlyGlnLeuGlyLysSerLysGlyAspLysGlyGly 1174

                                                              550 GATGAAGGGGCCGGCGGGCCCCATAGCCGTCCACGCCAGTGTCACTGGCTG 501
                                                                       :::|| |||:::     ||| ||||||:::||||
                                                              1175 GluGlnGlyProProGly..ProThrGlyPro...GlnGlyProIleGlyG 1190

                                                              500 AGCCT...AGGAGGCGGACACAGACCAGGCCCAGCACTGGACCAATGCCC 454
```

```
DR   PFAM; PF01410; COLFI; 1.
DR   PFAM; PF01391; Collagen; 17.
SQ   SEQUENCE   1840 AA;  184175 MW;  44D7E961 CRC32;

alignment_scores:
                 Quality:   334.50        Length:   1216
                   Ratio:    0.691          Gaps:     78
Percent Similarity:  39.803   Percent Identity: 24.836

alignment_block:
US-09-030-606-110/rev x Q60467

Align seg 1/1   to: Q60467   from: 1  to: 1840

3068 GGTCCTGCAGTAGCTCCAAACAGGGTTGTGGAGCTGGTGGGGAAAGTTGG 3019
     |||||||||:::||:::       :::||       ||| |      ||
 433 GlyProGlyMetProAlaAsnGlnAspThrIleTyrGluGlyLyIleGlyGl 449

3018 GGGTAGGGGAAAGTTGGGG...GTAGGGGAAATTTGGGCAGTGCCTTCA 2972
     | ||||||| |||||||||       ||||||||       |||||
 449 yProArgGlyGluLysGlyGlnLysGlyGlu........ProAlaIleI 463

2971 TCAGCCCAGTCCTAGAGAGAGTAGAGGG.................... 2943
     ||:::||| ::  ::|||||
 463 leGluProGlyMet...LeuIleGluGlyLysProProGlyProGluGlyPro 478

2942 .......................AGTGAAGTGGGGGAACCAGGC 2920
                            ||::|||    :::::
 479 AlaGlyLeuProGlyProProGlyThrProGlyThrGlyGlnValGl 495

2919 TGGGCCAAGAGAGAGAGGGGTGGTTAGGGAAGCCGTTGAGACCTGAAGCC 2871
     : ||| |||:::|||   ||||||   |||:::|||
 495 yAspProGlyGluArgGlyProProGlyArgProGlyLeuProGlyAlaA 512

2870 ................CCACCCTACCTTCCTCAACACCCTAACCTTGGG 2835
                     ||||::||||::||:::||:|||   |||
 512 spGlyLeuProGlyProProGlyThrMetLeuMetLeuPro......... 525

2834 TAACAGCATTTGGAATTATCATTTGGATGAGTAGAATTCCAAGGTCCT 2785
                         :::|||        ||:||:  ::
 526 ...............................PheArgPheGly...... 529

2784 GGGTTAGGCATTTGGGGGGCCAGA.................CCCCAGGAGA 2750
     ||| ::|                |||::|||        |||||:
 530 .GlyGlyGlyAspAlaGlySerLysGlyProMetValSerAlaGlnGlus 546

2749 AGAAGATTCTGGCAATGATCAGCCCAATGACCAGCTATCTCAGGGGA... 2703
     :::|| |||||||       ||||||||       |||||||
 546 erGlnAlaGlnAlaIleLeuGlnIleAlaArgLeuAlaLeuArgGlyPro 562

2702 ......................CCTGATTGTTGGGGATCCCCCAC 2680
                           |||   :| ||||||
 563 AlaGlyProMetGlyLeuThrGlyProGlyArgProGlyProMetGly.ProProG 579

2679 CCTACCCAAATATTAGCACCAACACAGAAAAGCTAGCAATGGATTCCCT 2630
     :::::::::|||:::||       ::       |||:::
 579 lySerGlyGlyLeuLysGlyGluProGlyAspMetGlyProGlnGlyPro 595

2629 TCTACTTTGTTAAATAAATAAAGTAAATATTTAAATGCCTGTCTCTGT 2580
                                             :::|||
 596 Arg...........................GlyVa 598

2579 GATGGCAACAGAGGACCAACAGGCCACATCCTGATAAAAGTAAGAGGG 2530
     |:::::     ||||||||||       ::  :::|||||||
 598 lGlnGlyProProGlyProThrGlyLysProGlyArgArgGlyArgAlaG 615

2529 GGGTGGATCAGCAAAAAGACAGTGCTGTGTGGGCTGGAGGGGACCTGGTTCTT 2480
     || |||       :::       |||
 615 lySerAsp.................. 617

2479 GTGTGTTGCCCCTCAGGACTCTTCCCCTACAAATAAGTCATATGTTCAAA 2430
```

(left column, bottom portion)

```
1190 InProGlyProSerGlyAlaAspGlyGluPro.....GlyProArgGlyGly 1204
     ::||| ||||||:::::       :::|||        ||||
 453 AGCACCATAGGTCATGAACTTCCCTACCCCCACTTCAGCAGCAGAGG 404
     ::  ||:::|||       ||:::||||||
1205 GlnIleGlyLeu.......PheGlyGlnLysGlyAspGluGlySerArgGl 1219

 403 CGGCACATAGTGATGCCTGGGCCAAACACACCTCCAGGC...CAAAGG 357
     |     ::|||:::|||  |||   |||  |||
1219 y.........PheProGly.ProProGlyProValGlyLeuGlnGly 1231

 356 TTAGCAGGTTGACACGACCAAGAGCTGGGCTTTCCCGT................ 321
     ||| |||  |||:::|||  |||  :::||
1232 LeuProGlyProProGlyGluLysGlyGluThrGlyAspValGlyGlnMe 1248

 320 ..........GCCCAGCAGGCGGCTCACCCACAG 296
                      ||  :::||||||
1248 tGlyLyProGlyProProGlyProArgGlyProSerGlyAlaProGlyA 1265

 295 CCCTGGACCATAGTGGGCCA....................... 275
     ||:::||||| ::||||||
1265 laAspGlyPro.GlnGlyProProGlyGlyIleGlyAsnProGlyAlaVa 1281

 274 .GGCGGGTAGGGCGTCAGGGGGCCGTTCAGGCACTCCAGAACTGCTTCGTC 226
     ||| :||  ||:::|||  ||| |||  :|||||
1281 lGlyGluLysGlyGlyLuProGlyGlyAlaGlyLyAspProGlyLeuProGlyG 1298

 225 TCGGC........................TCT 218
     ||                    |||
1298 luGlyGlyProLeuGlyProLysGlyGluArgGlyGlyGluLysGlyGluAla 1314

 217 GCTCCAGAAGCTGCGGCCTTCCTCCTTCGTCCGCGC............ 182
     ::|||:::||||||:::|||||||  :::
1315 GlyProSerGlyAlaAlaGlyGlyProProGlyProLysGlyProProGlyAs 1331

 181 .CAACTGCCTAGGAATCAGCCAGGCGCCCATTTCTGCCAGCCCTTTGT. 134
     ::  ||:::|||:::|||||  |||::|||
1331 pAspGlyProLysGlySerProGly................ProValGlyP 1343

 133 ........GCCGGTCCAGCTTCTCAGCCCATGCTCCAACCACCTGCTGC 95
     ||||| :::||||||       :::  |||||:::
1343 heProGlyAspProProGlyProGlyGluPro.......Gly.ProAlaGl 1357

  94 TGTGGGGCACCTCAGTGGGGACACGTCTCATCACTCAGATCCTGGCCCAG 45
     ||| ::|       :::||:::||||||
1357 yGlnAspGlyProProGlyAspLysGlyAspAspGlyProGlyGlnT 1374

  44 GCGGGCGGCTGTCACCCCGA 25
     ::     |||||||||
1374 hrGly......SerProGly 1378
```

```
seq_name: sp_rodent:Q60467

seq_documentation_block:
ID   Q60467       PRELIMINARY;      PRT;   1840 AA.
AC   Q60467;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PRO-ALPHA-1 TYPE V COLLAGEN.
DE   PRO-ALPHA-1
GN   PRO-ALPHA-1
OS   Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92105142.
RA   GREENSPAN D.S., CHENG W., HOFFMAN G.G.;
RT   "The pro-alpha 1(V) collagen chain. Complete primary structure,
RT   distribution of expression, and comparison with the pro-alpha 1(XI)
RT   collagen chain.";
RL   J. Biol. Chem. 266:24727-24733(1991).
DR   EMBL; M76730; AAA37002.1; -.
```

us-09-030-606-110.rspt

```
617  ..........................................................  617
2429 TCCCATGGGAGGAGTGTTTCATCCTAGAAACTCCCATGCCAAGCAAGAGCTACATT 2380
617  ..........................................................  617
2379 AAACGAAGCTGCAGGTTAAGGGGCTTAGAGATGGGAAAACCAGGTGACTGA 2330
618  ...............GlyAlaArg...GlyMetProGlyGln...  625
2329 GTTTATTCAGCTCCCAAAAACCCTTCTCTAGGTGTGTCTCAACTAGGAGG 2280
626  .....ThrGlyLysGly...AspArgGlyPheAspGlyLeuAlaGl 639
2279 CTAGCTGTTAACCCTGAGCCTGGGTAATCCACCTGCAGAGTCCCCGCATT 2230
639  y..............LeuProGlyGluLysGlyHisA 647
2229 CCAGTGCATGGAGCCCTTCTGGCCCTCCCTATAAGTCCAGACTGAAACC 2180
647  rgGlyAspProGlyProSerGlyProPro..............  656
2179 CCCTTGGAAGGCCTCCAGTCAGGCAGCCCCTAGAGACTGGGAGAGAGG.. 2132
657  ...........GlyLeuProGlyAspAspGlyGluArgGl 666
2131 yAspAspGlyGluValGlyProArgGlyLeuProGlyGluProGlyProA 2104
666  ...........AGAGGGACGGCCCCCAGCCCCCAGCTGTGC 2104
2103 AG..........CTACGCCACCTGCAGCAGCAGGTGGCAGCAGCAGAGGCCCACAT 2060
683  rgGlyLeuLeuGlyProLysGlyProGlyProProGlyProPro...  698
2059 TACTTTGGCAGCAACAGAAACTGGGGCCCCAGCCCCGGCCCCCCATGGGGC 2010
699  ...........Gly.ValThrGlyMetAspGlyGlnProGly...ProLysGlyA 712
2009 TAACAGGA...GCGGGAGCTGGACCCAGTGAGGCAGGCCCTCCACCCC 1963
712  snValGlyProGlnGlyGlyGluProGlyProGlyGlnGlyAsnPro 728
1962 AATGTGCTGGAAGTTTCTACGCTGAGTATTGGCCAAGTCGCTCTTGTC 1913
729  ...........GlyProAsnGlyGlnGlyProGlnGlyPheProGlyProLysG 739
1912 AAATACTACCTGTGTAGCAAAGTAAATGGCAGACCCAGGCCTGCGG 1863
739  ...........AlaGlnGlyLeu.ProGlyProGlnGlyA 755
1862 CAGACACCATATAGGCAGTGACAGACTGGCTAGCTGGACAATGGAGCCC 1813
756  ...........GlyMetProGlyAlaAspProGlyProHisProGlyLys 769
1812 ATAAACAGGGATGGGGACCCTGGG...ACAGCAGGAAGCCACTATCCAG 1766
770  ...........GluGlyProProGlyGluLysGlyGlyGlnGlyProProG 783
1765 GATGGCGAGGTCCAGGCAGATGCCCCGGCCGGAA.............. 1731
783  lyProGlnGlyProIleGlyTyrProGlyProArgGlyValLysGlyAla 799
1731 .......................................... 1731
800  AspGlyIleArgGlyLeuLysGlyLysGlyThrLysGlyGluLysGlyGluAspGl 816
1731 .......................................... 1731
816  yPhePr oGlyLysGlyAspMetGlyIleLeuLysGlyAsnArgGlyGluI 833
1730 .:CCACCAGCAGTGTGGCTGCTGCTCACCCACCACCACTACGGAGA 1686
833  leGlyProProGlyProArg...GlyGlu 841
```

```
1685 CATCACAGGCAGAGGCCCCGCAGAGCCGGGTGGAGGTGGGAGCAGGCCA 1636
842  ...AspGlyProGluGluGlyAr gGlyArgGlyGlyProAsnGlyAspPro 857
1635 CTGCCTCCAGCACCACGTGTCCATTAGGGAAG...GGAGCTCCAGGCTT 1589
858  GlyProLeuGlyProThrGlyLeuLysGlyLeuGlyValProGlyLe 874
1588 AGGG......CCTGGCAGGAAGCTGGTCATCAGGCTGTCCTCACTGCTAG 1545
874  uProGlyTyrProGlyArgGln.............G 882
1544 CACCTCCAGTGTCCCCTCGGTATT......TGGGCAGGAACACCTGCTTC 1501
882  lyProLysGlySer.IleGlyPheProGlyAlaAsnGlyGl 898
1500 TCCCGGTGGTAGAGGAGGCCAGTGTGTAGGGCAGGATCTGCAGGGCTGA 1451
898  uLysGlyGlyArgGlyThrPro.............GlyLysProGly.... 909
1450 GAAGGTGAACCCCGTGAGGGGGCTGAAGCTGTCACCACGGCCACACTGT 1401
909  .......................................... 909
1400 GGGACAGGCATGTGGCACCGGCAGCCCACCAGGGAAAGCTGCCACACTGGCC 1351
910  .......................................ProArgGlyGln. 913
1350 AAATAGACTGCTTCGAGTGCCGAATCGCTGCACCAGCCGGTCCATGACCAG 1301
914  ................................ArgGlyProThrGlyProArgGlyG 922
1300 AGAGA..................AGACCAGG...AGATGGCGCACT 1275
922  luArgGlyProArgGlyIleThrGlyLysProGlyProLysGlyAsnSer 938
1274 GCAGGAACAGCCCA...GGCTGCCATCCGAACGCCTTCATCATAGTGT 1228
939  GlyGlyAspGlyProAlaGlyProProGlyGluArg............. 950
1227 CTCCGGGCCCTCGGTGCCCGGCCTCAGCTCTGGGCACGCCCCGGTACAGCCCC 1178
951  ....GlyProAsnGlyProGlnGlyProThrGlyPheProGlyProLysG 966
1177 CTCGCCCCACGAAATCCGTGTAAAAACAGCGTGAAGGTCATGAGTGCCA.. 1131
966  lyProProGlyProProGlyLysAspGly...........LeuProGly 978
1130 .....................TCCAGCTGCACAGCTCAGCC 1111
979  HisProGlyGlnArgGlyGluGlyThrGlyPheGlnGlyLysThrGlyProPr 995
1110 ACGAAGAGCCCGGCCAGAGGGTGCGGGGCATGGGCAGCACA......GCTG 1067
995  oGlyProProGlyVal...ValGlyProGlnGlyProThrGlyGluThrG 1011
1066 GTGCACGCCGGGGAAGCAGCAGGGCCGGTTCCGGAAAGCCAAGCGGG... 1020
1011 lyProMetGlyLeuArgGlyHisProGlyProProGlyProGlyGlu 1027
1019 .......CCCGGCATGGACAGCAGCAGTGGGGCGGCACAAGGAGGGGGCCGAC 979
1028 GlnGlyLeuProGlyValAlaGlyThrLysGlyGluAspPro.... 1043
978  AGCCCTTCTGCCTCGGTCGGGTGGGCCCCAGCG...CTGCCCCTCAG.... 936
1044 ...................GlyProAlaGlyLeuProGlyLysLysAspG 1053
935  .:CCACCAGCAGTGTGGCTGCTGCTCACCCACCACCACTACGGAGA 888
1053 ..............GlyLeuArgGlyPhePro 1061
```

Wed Sep 29 14:26:57 1999

```
1330 oProGlyAspAspGlyProLysGlySerProGly.................P 1342
              ||| :::::::::::::|||||
          ..GCCGGTCCAGCTTCTCAGCCATGCTCAAC 104
141  CCTTTGGT.........                                 ::
     || ||| ||||||
1342 roValGlyPheProGlyAspProGlyProProGlyGluPro.....Gly 1356
     ||||||:|||             ::::
103  ACCTGCTGCTGTGGGCACCTCAGTGGGACACGTCTCATCACTCAGATC 54
     :||||||:::|          |||||||    ::: :::|||||||
1357 .ProAlaGlyGlnAspGlyProProGlyAspLysGlyAspAspGlyGluP 1373
53   CTGGCCGAGGCGCGGCGGCTGTCACCCGGA 25
     |||||: :::    ||||||||
1373 roGlyGlnThrGly......SerProGly 1380
```

seq_name: sp_plant:Q9ZTB9

```
seq_documentation_block:
ID   Q9ZTB9      PRELIMINARY;       PRT;    512 AA.
AC   Q9ZTB9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   SUCROSE TRANSPORTER.
GN   SUT1.
OS   Apium graveolens (Celery).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Apium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LEAF;
RA   NOIRAUD N., DELROT S., LEMOINE R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF063400; AAC99332.1; -.
SQ   SEQUENCE 512 AA;  54520 MW;  D4C8FBDF CRC32;

alignment_scores:
           Quality:  333.00                Length:   555
           Ratio:    1.168                 Gaps:     21
Percent Similarity: 51.351     Percent Identity: 25.045

alignment_block:
US-09-030-606-110 x Q9ZTB9

Align seg 1/1  to: Q9ZTB9  from: 1  to: 512

332  CAGCTCTTGCTGTCAACCTGCTCAACCTTGGCCTGGAGGTGTGTTTGGC 381
     |||||||:|||           |||:::|||||
33   LysLeuIleLeuValAlaAlaIleAlaAlaGlyValGlnPheGlyTrpAl 49
382  CCGAGGCATCACCTATGTGCCGCCCTCGTCTGCTGGGAAGTGGGGTAGAGG 431
     :::::::|||            |||             ::::|||||
49   aLeuGlnLeuSerLeuLeuThrProTyrValGlnLeuLeuGlyIleProH 66
432  AGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTTGCTCCGGCCTGGTC 481
     :||||:||| |||     ||||||    ||||||||
66   isLysTrpAlaAlaTyrIleTrpLeuCysGlyProIleSerGlyMetLeu 82
482  TGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCACGCCATGG 531
     ||||||:::|||       |||||||    ||||||||
83   ValGlnProIleValGlyTyrTyrSerAspArgCysGlnSerSerPheGl 99
532  CCGCCGCCGCCCTCCATCTGGCACTGCTCCTTGGCATCCTGCTGAGCC 581
     ||||||||||||||| ::: :||||||    :::||||
99   yArgArgArgProPheIleAlaSerGlyAlaGlyGlyCysValAlaIleSerV 116
582  TCTTTCTCATCCCAAGGGCC.........GGCTGGCTAGCAGGGCTGCTG 622
     :::||||||||||:::|||             |||||
116  alIleLeuIleGlyPheAlaAlaAspIleLeuTyrLysAlaGly.......130
623  TGGCCGGATCCC...........AGGCCCCTGGAGCTGGCCACTGCT 657
```

```
887  GGCCAAGAGGCACT.................... 873
     ||||||:|||||
1062 GlyAspArgGlyLeuProGlyProValGlyAlaAlaLeuGlyLeuLysGlySe 1078
872  ......CCTCCTGGGTGCCCAGGT......AGGGGCCAGGCACTGGT 836
     |||||||    ||||||      :|||||:|||
1078 rGluGlyProGlyProProGlyProAlaGlySerProGlyGluArgG 1095
835  GTCCCAGTCAATGGCAGCAGGAGGTAGCCCAGGCAGCCCCCAAGAACTGA 786
     ||||::::      |||::::|||     |||  |||
1095 lyProAla......GlyAlaAlaGlyProIleGlyIlePro......... 1106
785  TCATGAAGGCATAGACAGAGTAGGCCTGGCGACAGTGTCCGGGTCCCGG 736
     :::||||||||      ::|||||:|||||||||
1107 ................GlyArgProGlyProGlnGlyPro......... 1114
735  AAGAGGTCAGAGAGCAGGGCCTCCAGTGAGTGAAGCACACTGGCCACA 686
     ||||||||||   |||     :::||||||:|||:
1115 .........ProGlyProAlaGlyGluGlyValProGlyValGluL 1127
685  GAAGTCCAGCAGCCCACGCCCAGGATGAGCAGTGCCAGCTCCAGGGGCC 636
     ::::||| :::||| ||    |||||||||||
1127 ysGlyProGlnGlyTyrPro.AlaGlyArgAsp.....GlyLeuGlnGlyPr 1141
635  TGGGATCCGGGACAGCAGCCCTGCTAGCCAGCCGGCCGTTGGGATGAG. 587
     |        |||:::|||||    ::::|||
1141 o...ValGlyLeuProGlyProAlaGlyProValGlyProProGlyGluA 1157
586  .............AAAGAGGCTCAGCAGGATGCCCAAGGA 560
     ::  ||||||:|||     :::|||||
1157 spGlyAspLysGlyGluIleGlyGlyProGlyGlnLysGlySerLysGly 1173
559  CAGTGCCCAGATGAAGGGCCGGCGGCCCATAGCGTCCACGCCCAGTGGT 510
     :::         ||||||| ::: |||  ||| |||||||||
1174 AspLysGlyGluGlnGlyProProGly.ProThrGlyP...GlnGlyP 1189
509  CACTGGCTGAGCCT...AGGGACGGGACAGCAGCAGGCCCAGCACTGGA. 463
     |||||||||        |||||||    ::::|||
1189 roIleGlyGlnProGlyProSerGlyAlaAlaSpGlyProGlyGluPro......Gly 1203
462  CCAATGCCCAGCACCATGGTCATGAACTTCTCCTCTACCCCCACTTCCAG 413
     |||  ||| ||| :: ||| :::::|||||
1204 ProArgGlyGlnGlnGlyLeu.....PheGlyGlnLysGlyAspGlyGluGl 1218
412  CAGCAGAGGCGGCACATAGGTGATGCCTGCGGCCAAACACACTCCAGGC 363
     :||||||||||      |||        ||| |||
1218 ySerArgGly.......PheProGly. ProProGlyProValGlyLy 1230
362  ..CAAAGGTTAGCAGTTGACCAGCAAGAGCTGGGCTTTCCGGT...... 321
     || ||| ||| |||:::|| ||||| :||||
1231 LeuGlnGlyLeuProGlyProGlyProGlyGluLysGlyGluThrGlyAspVa 1247
320  ...........GCCGCCAGGCGGCT 305
     ::|||
1247 lGlyGlnMetGlyPro......GlyGlyProArgGlyProArgGlyGlyA 1264
304  CACCCCAGCGCCTCTGACCACCATAGTGGGCCA................. 275
     :|||  ||||:::||| |||
1264 laProGlyAlaAspGlyPro.GlnGlyProProGlyLyIleGlyAsnPr 1280
274  .........GGCGGGTAGGGCTCAGGGGCCGTTCAGGCACTCCAGAAC 235
     ||         ||||:::|| |||||||||
1280 oGlyAlaValGlyLuLysGlyLyGluProGlyGluAlaGlyGluProGlyL 1297
234  TGCTTCGTTCGGCTCT..................... 218
     || |||:::
1297 euProGlyGluGlyGlyProLeuGlyProLysGlyGluArgGlyGluLys 1313
217  .........GCTTCCAGAAGCTCGCGGCCTCTCCTTGCTGCCGCC..... 182
     ::|||||||||:::|||||
1313 yArgArgArgProPheIleAlaSerGlyAlaGlyGlyCysValAlaIleSerV 1330  (see above)
314  GlyGluValGlyLyProSerGlyAlaAlaLaGlyProSerGlyAlaGly......... 130
181  ...:?.....CAACTGCTAGGAATCAGCCAGGCGCGCCCATTTCTGCCAGC 142
```

```
131  ......AspAspMetAsnLysThrLeuLysProArgAlaValThrValPh  145
           :::::::::::|||
658  CATCCTGGGCTGGGCTGGACTTCTGTGGCCAGGTGTGCTTCACTC       707
     ::::::|||      |||      |||
145  evalIleGlyPheTrpIleLeuAspValAlaAsnAsnMetLeuGlnGlyP  162
     ||  :::|||||||||||||  ::: |||
708  CACTGGAGGCCCTGCTTCTCGACCTCTTCCGG...GACCCGGACCACTGT  754
                                       :::
162  roCysArgAlaLeuLeuAlaAspLeuCysAsnGlyAspThrArgArgMet  178
775  CGCCAGCCTACTCGTCTATGCCTTCATGGTCTTGGGGCTGCCT          804
     |||  ::::::|||  |||||||||::|||  |||||||||  ||
179  ArgSerAlaAsnAlaLeuTyrArgPheMetAlaValGlyAsnIleLe     195
805  GGGC                                               
     |||                        TACCTCCTGCCTGCCATTG
195  uGlyAsnAlaAlaGlySerTyrAsnAsnLeuTyrLysLeuProPheS     212
     :::                     |||  ::: |||
828  ACTGGGACACCAGTGCCCTGGCCCCTACCTGGCCACCAGGAGGACTGC    877
     ::  |||  |||  :::::: |||::::::|||
212  er...LysThrHisAlaCysAspLeuTyrCysAlaAsnLeuLysSerCys  227
878  ...CTCTTTGGCCTGCTCACCCTCATCTTCCTCACCTGCGTAGCAGCCAC  924
     :::::::::::|||  ||||||||||||  ||||  |||::::::|||
228  PheIleIleSerIleValLeuLeuIlePheIleThrValLeuAlaLeuTh  244
925  ACTGCTGGTGGCTGGAGGAGGCGGCTGGGCCCCACCGAGCCAGCAGAAG   974
     |||:::::|||  |||::||  |||
244  r......ValValArgGlyLysGlnTrpSerProAspGluAlaAspGluG  259
975  GGCGTCGGGCCCCCTCCTGTGCCCCACTGCTGTCCATGCCGGGGCCCGC  1024
     |||          :::  :::|||
259  luProProSerGlyLysIlePro                             267
1025 TTGGCTTCCGGAACCTGGGCGCCCTGCTTCCCCGGCTGCACCAGCTGTG  1074
     :::::::||||||||||  :::::|||  ::|||
268  ..........ValPheGlyGluLeuLeuArgAlaLeuLysAspLeu..    279
1075 CTGCCCCATGCCGCACCCCGCCGCCGCTCTTCGTGGCTGAGCTGTGCA   1124
     |||||        |||::::||
280  ..........ProArgProMetLeuMetLeuLeuLeuAlaValThrCysLeuA  293
1125 GCTGGATGGCACTCATGACCTTCACGCTGTTTACACGGATTCGTGGGC   1174
     :::||:::|||  ::: |||  ||||||  ::::::|||
293  snTrpIleAlaTrpPhePheProPheIleLeuPheAspThrArgTrpMetGly  309
1175 GAGGGGCTGTACCAGGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCG  1224
     ::::::  ||||||||
310  ArgGluIleTyrGlyGlyGlyThr......AlaGlyGlnGly..........  320
     :::::::|||  |||    |||
1225 GAGAACTATGATGAAGGCGGTTCGGATGGGCCAGCCTGGGCGTGTTCCTGC  1274
     :::  ||||||||||||||||:::||||||||::|||
321  .LysLeuTyrAspGlnGlyValArgValGlySerLeuGlyLeuLeuLeuA  337
1275 AGTGCGCCATCTCCCTGGTCTTCTCTGGTCATGGACCGGCTGGTGCAG   1324
     ::  :::::::  :::  :::::::        |||||||:::|||
337  snSerValLeuGlyLeuThrSerIleLeuAlaValGluTyrLeuValArg  353
1325 CGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTTCCCTGT  1374
     |||      |||  ::: |||
354  GlyValGlyValLysIle...LeuTrpGlyLeuValAsnPheLeuLe     369
1375 GGCTGGCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGTGACAGCTT  1424
     ::  ::::::|||  |||||||:::|||||::|||
369  uAlaIleGlyLeuVal..........MetThrValValSerL          381
1425 CAGCCGCCCTCACCGGGTTCACCTTCTCAGCCTCGAGATCCTGCCCTAC   1474
     :::::::|||
381  ysValAla..........                                  383
1475 ACACTGGCCTCCCCTCCTACCACCGGGAGAAGCAGGTGTTCCTGCCCAAATA  1524
     |||::::::::::
```

```
384  ......GlnHisGlnArgGlnHis......                      389
1525 CCGAGGGGACACTGGAGGTGCTAGCAGTGAGCAGCCTGATGACCAGCT   1574
389  ......                                              389
1575 TCCTGCCAGGCCCTAAGCCTGGAGCTCCTTCCCTAATGGACACTGGGT   1624
389  ......                                              389
1625 GCTGGAGGCAGTGGCCTGCTCCACCTCGCGCTCTGCGGGGCCTC        1674
     |||:::::|||  |||  |||||  ::|||
390  ...GlyAlaAsnGlyGlnLeuLeuProProSerAla.....GlyValLy  403
     :::|||||||||||||||
1675 TGCCTGTGATGTCTCCTACGTGGTGGTGGTGGTGAGCCCACCGAGGCCA  1724
     :::      ::: |||
403  sAlaGlyAlaLeuSerLeuProPheSerIleLeuGlyIleLeuProLeuSerIleT  420
1725 GGGTG..GTTCCG......                              GGC  1738
     ::::||
420  hrPheSerIleProPheAlaLeuAlaSerIleTyrSerSerGlySerGly  436
     ::|||
1739 CGGGGCATCTGCCTGGACCTCGCCATCCTGGATAGTGCCTTCCTGCTGTC  1788
     |||  :::::|||:::|||::::::|||||::: |||
437  AlaAlaGlyGlnGlyLeuSerLeuGlyValLeuValLeuAsnLeuAlaIleValValPr  453
1789 CCAGGTGCCCCATCCTGTTTATGGGCTCCATTGTCCAGCTC......A  1832
     :::  :::|||  |||::::|||
453  oGlnMetIleValSerValLeuLeuAlaGlyProPheAspSerLeuPheGlyG  470
1833 GCCAGTCTGTCACTGCCTATATCGGTG.......TCTGCCCGCCAGGCCTG  1873
     :::::::::::|||||||||||
470  lyGlyAsnLeuProAlaPheValValGlyAlaIleAlaIleSerAlaAlaIleSer  486
1874 GGTCTGGTCGCCATT                                   1888
     |||::::::|||||
487  GlyValLeuAlaIle                                    491
```

seq_name: sp_plant:Q43653

seq_documentation_block:

alignment_scores:
Quality: 333.00   Length: 539
Ratio: 1.198   Gaps: 21
Percent Similarity: 51.577   Percent Identity: 24.675

alignment_block:

us-09-030-606-110.rspt

US-09-030-606-110   to: Q43653

Align seg 1/1   to: Q43653   from: 1   to: 516

```
278 CCCACTATGGTCCAGAGGCTGTGGGTGAGCCGCCTGCTGCGGCACCGGAA 327
    |||  :::::|::|||||||
 24 ProLeuAlaProSerLysLeuTrp......................... 31

328 AGCCCAGCTCTTGCTGGTCAACCTGCTAACCTTTGGCCTGGAGGTGTGTT 377
    :::::||||||||   :::::|||   ||||:::::
 32 ...LysIleIleValValAlaSerIleAlaAlaGlyAlaValGlnPheGlyT 47

378 TGGCCGCAGGCATCACCTATGTGCCGCTCTGCTGCTGGAAGTGGGGTA 427
    ||   |||||   ||||||||   ::||:||   |:||
 47 rpAlaLeuGlnLeuSerLeuLeuThrProTyrValGlnLeuLeuGlyIle 63

428 GAGGAGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGCTGGGCCT 477
    ::||||||   ||||||:   ||||||:::   |||:|
 64 ProHisLysPheAlaSerPheIleTrpLeuCysGlyProIleSerGlyMe 80

478 GGTCTGTGTCCCGTCCTAGGCTCCAGCCAGTGACCACTGGCGTGGACGCT 527
    :::   ||||:::|||   |||||||   ::||||
 80 tIleValGlnProValValGlyTyrTyrSerAspAsnCysSerSerArgP 97

528 ATGCCGCCGCCGCCCTTCATCTGGCACTGTCCTTGGCCATCCTGCTG 577
    :::|||||||||||||:||   :::|||   ||||||
 97 heGlyArgArgProPheIleAlaAlaAlaGlyAlaAlaLeuValMetIle 113

578 AGCCTCTTTCTCATCCCCAAGGGCCGGCTGGCTAGCAGGGCTG 619
    :::|||||||   |||   |||
114 AlaValPheLeuIle.......GlyPheAlaAlaAspLeuGlyHisAl 127

620 ..........CTGTGCCCGGATCCCAGGCCCCTGGAGCTGGCACTGC 656
             |||   ::||||:::|||   ::|||||
127 aSerGlyAspThrLeuGlyLysGlyLyPheLysProArgAlaIleAlaValP 144

657 TCATCCTGGGCGTGGGGCTCTGACTTCTGTGGCCAGTGTGCTTCACT 706
    :::::||||   ::||||||   |||:||||||:
144 heValValGlyPheTrpIleLeuAspValAlaAsnMetLeuGlnGly 160

707 CCACTGGAGGCCCTGCTCTCTGACCCTCTTCCGGGACCCGGAC...CACTG 753
    ||   ::||||||||||::::|||||||   :::
161 ProCysArgAlaLeuLeuAlaAspLeuSerGlyLysSerGlyArgMe 177

754 TCGCCAGGCCTACTCTGTCTATGCCTCTCATGATCAGTCTTGGGGGCTGCC 803
    |||   ||   :::::||||   :::||::|||||   |
177 tArgThrAlaAsnAlaPhePheSerPheMetAlaValGlyAsnIleL 194

804 TGGGCTACCTCCTGCCTGCCCATTGACTGG................. 832
    ||||||   :::||   ||||||
194 euGlyTyrAlaLeuAlaGlySerTyrSerHisLeuPheLysValPhePhe 210

833 ...GACACCAGTGCCCTGGCCCTACCTGGCACCCAGGAGGAGTGCCT 879
    |||:||   :::::||||:   ::||||||
211 SerLysThrLysAlaCysAsnAlaPheTyrCysAlaAsnLeuLysSerCysPh 227

880 CTTT...GGCCTGCTCACCCTCATCTTCCTCACCTGCCTAGCAGCCACAC 926
    ||||   :::::|||   |||||||   |||
227 ePheIleAlaIlePheLeuLeuLeuSerLeuThrThrIleAlaLeuThrL 244

927 TGCTGTGGCTGAGGAG...GCAGCGCTGGGCCCCACCGAGCCAGCAGAA 973
    ||:::|||   |||   :::::|||
244 euValArgGluAsnGluLeuProGluLysGluGlnGlnIleAspGlu 260

974 GGGGCTGTCGGCCCCCTCCTTGTCGCCCACTGCTGTCCATGCCGGGCCCG 1023
    ||||:::::   :::
261 LysLeuLeuAlaGlyAlaGly.......................LysSerLy 269

1024 CTTGGCTTCCGGAACCTGGGCGCCCCTCGCCGGCTGCACCAGTCGT 1073
    :|||   |||   :::::||||   |||   :::::|||
269 sValProPhe.....PheGlyLeuIlePheGlyAlaLeuLysGluLeu. 283

1074 GCTGCCCATGCCCCACCCCTGCGCCGGCTCTTCGTGGCTGAGCTGTGC 1123
```

```
284 .........ProArgProMetTrpIleLeuLeuLeuValThrCysLeu 296
          |||||   ::   ::::::
1124 AGCTGGATGGCACTCAGTGACCTTCACGCTGTTTACACGGATTTCGTGGG 1173
     ::|||:::||   :::   |||   |||:   |||:||:::|:::
1297 AsnTrpIleAlaTrpPheProPheLeuTyrArgAspThrAspTrpMetAl 313

1174 CGAGGGGCTGTACCAGGGCGTGCCCAGCTCCGGCACCGGCACCGAGGCCC 1223
     ::::   ::||   |||   |||
313 aLysGluValPheGlyGly...........GlnValGlyAsp......A 324

1224 GGAGAACTATGATGAAGGCGTTCGGATGGCCAGCCTGGGCTGTTCCTG 1273
     ||   ||||||   ||||||||   ||:::::||||:::|||
324 lArgLeuTyrAspLeuGlyValArgAlaGlyAlaMetGlyLeuLeuLeu 340

1274 CAGTGCGCCATCTCCCTGGTCTTCTCTCTGGTCATGGACCGGCTGGTGCA 1323
     |||   ::::||   ||   :::||::::   |||   |||:::
341 GlnSerValValLeuGlyPheMetSerLeuGlyValGlyGluPheLeuGlyLyLy 357

1324 GCGATTCGGCACTCGAGCAGTCTATTGCCCAGTGTGCAGCTTCCCTG 1373
     :::::::|||   |||   :::::|||   ::   :
357 sLysIleGly...GlyAlaLysArgLeuTrpGlyIleLeuAsnPheVaIL 373

1374 TGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGTG........ 1417
     ::   ||||:::   :::::|||:
373 eu.........AlaIleCysLeuAlaMetThrIleLeuValThrLysMet 386

1418 ...............ACAGCTTCAGCCGCCCTCACCGGTCAC 1446
                 |||   |||   ||   ||   ||
387 AlaGluLysSerArgGlnHisAspProAlaGlyThrLeuMetGlyProTh 403

1447 C..................TTCTCAGCCCTGCAGATC. 1465
                     |||::||||||   |||
403 rProGlyValLysIleGlyAlaLeuLeuPheAlaAlaLeuGlyIleP 420

1466 ...............CTGCCCTACACACTGGCCTCCCTCCTACCAC 1495
                 ::|||||:||||||:|||||||:::
420 roLeuAlaLaThrPheSerIleProPheAlaLeuAlaSerIlePheSer 436

1496 CGGGAGAAG..........................CA 1506
     |||||||||   :::   ||||||||
437 SerAsnArgGlySerGlyLnGlyLeuSerLeuGlyValLeuAsnLeuAl 453

1507 GGTGTTCCTGCCCAAATACGAGGGGACACTGGAGGTGCTAGCAGTGAGG 1556
     :::   ::||:||   ||||||   |||:::   :::|
453 aIleValProGlnMetLeuValSerLeuValGlyGlyProTrpAspA 470

1557 ACAGCCTGATGACCAGCTTCCTGCCAGGCCCTAAGCCTGGAGCTCCCTTC 1606
     ::||:::::||||||   ||||||||   ||
470 spLeuPheGlyGlyAsnLeuProGly.................Phe 480

1607 CCTAATGGACACGTGGGTGCTGGAGGCAGTGGC............CT 1641
     ||   ::||||||   ||||||   ::
481 ValValGlyValAlaValAlaAlaAlaSerAlaValLeuAlaLeuThrMe 497

1642 GCTCCACCTCCACCCGCGGCTGCTGCGGGGCCCTCGCCCTGTGATGCTCCG 1691
     :||||||   |||||||
497 tLeuProSerProProAla.................AspAlaLysP 507

1692 TACGTGTGGTGGTGGT 1708
     ||:::::|||
507 roAlaValAlaMetGly 512
```

This Page Blank (uspto)

```
OM nucleic - nucleic search, using sw model

                GenCore version 4.5
        Copyright (C) 1993 - 1998  Compugen Ltd.

Run on:      September 28, 1999, 10:58:39 ; Search time 1809.22 Seconds
                                         (without alignments)
                                         3717.812 Million cell updates/sec

Title:           US-09-030-606-110
Perfect score:   3410
Sequence:        1 GGGAACCAGCCTGCACCGC.........AAAAAAATAAAAAAAAAAAAAA 3410

Scoring table:   IDENTITY_NUC
Searched:        254578 seqs, 986266752 residues
Database :       EST:*
            1:  em_est1:*
            2:  em_est2:*
            3:  em_est3:*
            4:  em_est4:*
            5:  em_est5:*
            6:  em_est6:*
            7:  em_est7:*
            8:  em_est8:*
            9:  em_est9:*
           10:  em_est10:*
           11:  em_est11:*
           12:  em_est12:*
           13:  em_est13:*
           14:  em_est14:*
           15:  em_est15:*
           16:  em_est16:*
           17:  em_est17:*
           18:  em_est18:*
           19:  em_est19:*
           20:  gb_est1:*
           21:  gb_est2:*
           22:  gb_est3:*
           23:  gb_est4:*
           24:  gb_est5:*
           25:  gb_est6:*
           26:  gb_est7:*
           27:  gb_est8:*
           28:  gb_est9:*
           29:  gb_est10:*
           30:  gb_est11:*
           31:  gb_est12:*
           32:  gb_est13:*
           33:  gb_est14:*
           34:  gb_est15:*
           35:  gb_est16:*
           36:  gb_est17:*
           37:  gb_est18:*
           38:  gb_est19:*
           39:  gb_est20:*
           40:  gb_est21:*
           41:  gb_est22:*
           42:  gb_est23:*
           43:  gb_est24:*
           44:  gb_est25:*
           45:  gb_est26:*
           46:  gb_est27:*
           47:  gb_est28:*
           48:  gb_est29:*
           49:  gb_est30:*
           50:  gb_est31:*
           51:  gb_est32:*
           52:  em_est20:*
           53:  em_est21:*
           54:  em_est22:*
           55:  em_est23:*
           56:  em_est24:*
           57:  em_est25:*
           58:  em_est26:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
                                    SUMMARIES

                                %
Result                        Query
No.       Score   Match  Length  DB  ID          Description

C   1     456.8   13.4   491    36  AA631143   AA631143 nq75g08.s
C   2     449.2   13.2   469    50  AI703348   AI703348 wd93b09.x
C   3     366     10.7   388    35  AA579486   AA579486 nf33g07.s
C   4     333.8    9.8   348    40  AA984323   AA984323 am84a12.s
C   5     324.2    9.5   375    30  AA225106   AA225106 nc21d11.r
C   6     314      9.2   315    36  AA640153   AA640153 np28b03.s
C   7     311      9.1   313    48  AI587483   AI587483 tr51c10.x
C   8     306      9.0   306    36  AA631024   AA631024 nq76g11.s
C   9     300.2    8.8   674    47  AI525162   AI525162 promrna-9
C  10     295.6    8.7   346    25  N95796     N95796 zb66h02.s1
C  11     282.6    8.3   301    50  AI696721   AI696721 wc56d05.x
C  12     281.8    8.3   299    47  AI468280   AI468280 tg57a01.x
C  13     280.6    8.2   342    28  AA112573   AA112573 zm28c12.r
C  14     262.4    7.7   264    34  AA492342   AA492342 ng81d12.s
C  15     255.6    7.5   288    35  AA579735   AA579735 nf39g07.s
C  16     253.4    7.4   287    35  AA570251   AA570251 nf39d12.s
C  17     242      7.1   242    47  AI472447   AI472447 t177a04.x
C  18     228      6.7   228    28  AA112574   AA112574 zm28c12.s
C  19     221.4    6.5   239    26  W24907     W24907 zb66h02.r1
C  20     214.8    6.3   253    35  AA552457   AA552457 nj90f02.s
C  21     214      6.3   214    36  AA652651   AA652651 ns64c11.s
C  22     165.4    4.9   363    42  AU023209   AU023209 AU023209
C  23     162.8    4.8   242    36  AA647708   AA647708 vq77h04.s
C  24     160      4.7   173    36  AA652452   AA652452 ns64e11.s
C  25     142.4    4.2   163    35  AA579320   AA579320 nj36e12.s
C  26     123.6    3.6   686    48  AI598307   AI598307 EST250010
C  27     120      3.5   378    42  AI137795   AI137795 UI-R-C0-h
C  28     118      3.5   320    46  AA998873   AA998873 UI-R-C0-h
C  29     117.8    3.5   282    46  AA957294   AA957294 UI-R-E1-f
C  30     102.6    3.0   550    29  AA137485   AA137485 mq98h06.r
C  31     101.2    3.0   559    39  C88248     C88248 C88248 Mous
C  32     101.2    3.0   573    39  C88345     C88345 C88345 Mous
C  33     101.2    3.0   556    42  AU022931   AU022931 AU022931
C  34     101.2    3.0   589    42  AU023994   AU023994 AU023994
C  35     101.2    3.0   411    44  AU043413   AU043413 AU043413
C  36     101.2    3.0   438    44  AU043506   AU043506 AU043506
C  37     101.2    3.0   459    44  AU043764   AU043764 AU043764
C  38     101.2    3.0   462    44  AU043793   AU043793 AU043793
C  39     101.2    3.0   566    44  AU044322   AU044322 AU044322
C  40     101.2    3.0   360    44  AU044508   AU044508 AU044508
C  41     101.2    3.0   297    44  AU045544   AU045544 AU045544
C  42     100      2.9   349    39  C86407     C86407 C86407 Mous
C  43      99.2    2.9   457    42  AU024466   AU024466 AU024466
C  44      98.8    2.9   185    35  AA589036   AA589036 v163d01.r
C  45      98.6    2.9   285    47  AI536638   AI536638 to14e05.x
```

```
                                   ALIGNMENTS

RESULT  1
AA631143/c
LOCUS       AA631143       491 bp   mRNA              EST    31-OCT-1997
DEFINITION  nq75g08.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1158206 3',
            mRNA sequence.
ACCESSION   AA631143
NID         g2553754
VERSION     AA631143.1   GI:2553754
```

```
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 491)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1400896.

             Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html

             Insert Length: 919    Std Error: 0.00
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1..491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1158206"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT     135 a     104 c     136 g     116 t
ORIGIN
```

```
Query Match              13.4%;   Score 456.8;  DB 36;   Length 491;
Best Local Similarity   98.0%;   Pred. No. 8.8e-70;
Matches  484;  Conservative   0;  Mismatches   7;  Indels   3;  Gaps    2;

Qy   2837  CAAGGTTAGGGTGTTGAAGGAAGGGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCCTA  2896
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    491  CAAGGTTAGGGTGTTGAAGGAAGGTAGAGGGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCCTA  432

Qy   2897  ACCACCCCTCTCTTGGCCCAGCCTGGTTCCCCCCCACTTCCACTCCCCCTCACTCTCT  2956
           ||||||||||||||||    |||||||||||||||||||||||||||||||||||||||
Db    431  ACCACCCCTCTTTTCTTG--CCCAGCCTGTTCCCCCCTACTTCCACTCCCCCTCACTCTCT  374

Qy   2957  CTAGGACTGGCTGATGAAGGCACTGCCCAAAATTTCCCCTACCCCCCCAACTTTCCCCTAC  3016
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    373  CTAGGACTGGCTGATGAA-GACTGCCCAAAATTTCCCCTACCCCCCAACTTTCCCCTAC  315

Qy   3017  CCCAACTTTCCCCACCAGCTCCACAACCCTGTTGGAGCTACTGCAGGACCAGAAGCAC  3076
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314  CCCAACTTTCCCCACCAGCTCCACAACCCTGTTGGAGCTACTGCAGGACCAGAAGCAC  255

Qy   3077  AAAGTCGGGTTTCCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGG  3136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254  AAAGTCGGGTTTCCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGG  195

Qy   3137  AATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCA  3196
```

```
Db    194  AATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCA  135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   3197  GGGGGGTTTAAGTGCCGTTGCAATAATGTCGTCTTATTATTATTTAGCGGGGTGAATATT  3256
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134  GGGGGGTTTAAGTGCCGTTGCAATAATGTCGTCTTATTATTATTTAGCGGGGTGAATATT  75

Qy   3257  TTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTCT  3316
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74  TTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTCT  15

Qy   3317  TATATGTTTAAAAA  3330
           ||||||||||||||
Db     14  TATATGTTTAAAAA  1
```

```
RESULT 2
AI703348/c
LOCUS       AI703348             469 bp    mRNA        EST     03-JUN-1999
DEFINITION  wd93b09.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2339129 3',
            mRNA sequence.
ACCESSION   AI703348
NID         g4991248
VERSION     AI703348.1  GI:4991248
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 469)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188436.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1..469
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2339129"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "   4 others
BASE COUNT     126 a      97 c     135 g     107 t
ORIGIN
```

```
Query Match              13.2%;   Score 449.2;  DB 50;   Length 469;
Best Local Similarity   98.5%;   Pred. No. 1.8e-68;
```

```
Matches 462;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy   2857  AAGTAGAGGTGGGGCTTCAGGTCTCAACGGCGTTCCCTAACCACCCCTCTTCTTCTTGGC  2916
                |||||||||||||||||||||||||||||||||| |||||| |||||
Db    469  AAGTAGAGGTGGGGCTTCAGGTCTCAACGGCGTTCCCTAACCACCCCTCTTCTTCTTGGC   410

Qy   2917  CCAGCCTGGTTCCCCCACTTCCACTTCCACTTCCCCTACTCTCTAGGACTGGGCTGATGAAG  2976
                |||||||||||||||||||||||||| || |||||| || |||||
Db    409  CCAGCCTGGTTCCCCCACTTCCACTTCCACTCCCCTACTCTCTATCAGGACTGGNCTNATGAAG   350

Qy   2977  GCACTGCCCAAAATT-CCCTACCCCCAACTTTCCCCTACCCCCAACTTTCCCCACCAG  3035
                ||||||| ||||||  |||||||||||||||||||||||||||||||||
Db    349  GCACTGCNCAAAATTNNCCCTACCCCCAACTTTCCCCTACCCCCAACTTTCCCCACCAG   290

Qy   3036  CTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAGC  3095
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db    289  CTCCACAACCCTGTTTGGGAGCTACTGCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAGC   230

Qy   3096  CTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACCAGAAACTCA  3154
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db    229  CTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACCAGAAACTCA   170

Qy   3156  GGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATTCTCTCAGGGGGGTTTAAGTGCCGT  3214
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db    169  GGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATTCTCTCAGGGGGGTTTAAGTGCCGT   110

Qy   3216  TTGCAATAATGTCGTCTTATTATTTATTAGGCGGGTGAATATTTATACTGTAAGTGAGCAA  3274
                |||||||||||||||||||||||||||||||||||||||||||||
Db    109  TTGCAATAATGTCGTCTTATTATTTATTAGGCGGGTGAATATTTATACTGTAAGTGAGCAA    50

Qy   3276  TCAGAGTATAATGTTTATGTGACAAAATTAAAGGCTTTTCTTATATGTT  3324
                |||||||||||||||||||||||||||||||||||||||||||||
Db     49  TCAGAGTATAATGTTTATGTGACAAAATTAAAGGCTTTTCTTATATGTT     1
```

```
RESULT 3
AA579486/c
LOCUS       AA579486      388 bp    mRNA             EST      03-SEP-1997
DEFINITION  nf33g07.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:915612
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AA579486
NID         g2357670
VERSION     AA579486.1  GI:2357670
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On May 8, 1995 this sequence version replaced gi:800271.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
FEATURES             Location/Qualifiers
     source          1..388
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:915612"
                     /clone_lib="NCI_CGAP_Pr1"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected, histologically normal
                     prostate epithelial cells. Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies).  Average insert size is 600
                     bp. NOTE: Not directionally cloned.  This library was
                     constructed by David Krizman."
BASE COUNT      99 a     77 c     99 g    113 t
ORIGIN
```

```
Query Match             10.7%;  Score 366;  DB 35;  Length 388;
Best Local Similarity   98.7%;  Pred. No. 4e-54;
Matches 369;  Conservative   0;  Mismatches   5;  Indels   0;  Gaps   0;

Qy   2975  AGGCACTGCCCAAAATTCCCCTACCCCCAACTTTCCCCTACCCCCAACTTTCCCCACCA  3034
                ||||||||||||||| |||||||||||||||||||||||||||||||||
Db    381  AAGCACTGCCCAAAATTCCCCTACCCCCAACTTTCCCCTGCCCCCAACTTTCCCCACCA   322

Qy   3035  GCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAG  3094
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321  GCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAG   262

Qy   3095  CCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACCAGAAACTC  3154
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db    261  CCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACCAGAAACTC   202

Qy   3155  AGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATTCTCTCAGGGGGGTTTAAGTGCCG  3214
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db    201  AGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATTCTCTCAGGGGGGTTTAAGTGCCG   142

Qy   3215  TTTGCAATAATGTCGTCTTATTATTTATTAGGCGGGTGAATATTTATACTGTAAGTGAGCA  3274
                |||||||||||||||||||||||||||||||||||||||||||||||
Db    141  TTTGCAATAATGTCGTCTTATTATTTATTAGGCGGGTGAATATTTATACTGTAAGTGAGCA    82

Qy   3275  ATCAGAGTATAATGTTTATGTGACAAAATTAAAGGCTTTTCTTATATGTTTAAAAAAAAA  3334
                |||||||||||||||||||||||||||||||||||||||    |||||||||||
Db     81  ATCAGAGTATAATGTTTATGTGACAAAATTAAAGGCTTTCTTAACCGTTTAAAAAAAAA    22

Qy   3335  AAAAAAAAAAAAAA  3348
                ||||||||||
Db     21  AAAAAAAAAAAAAA     8
```

```
RESULT 4
AA984323/c
LOCUS       AA984323      348 bp    mRNA             EST      27-MAY-1998
DEFINITION  am84a12.s1 Stratagene schizo brain S11 Homo saplens cDNA clone
            IMAGE:1629790 3', mRNA sequence.
ACCESSION   AA984323
NID         g3162848
VERSION     AA984323.1  GI:3162848
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., Wolfe,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
```

```
JOURNAL      Unpublished (1997)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1402298.

             Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Possible reversed clone: polyT not found
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 348.
FEATURES     Location/Qualifiers
     source  1..348
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:1629790"
             /clone_lib="Stratagene schizo brain S11"
             /sex="male"
             /tissue_type="schizophrenic brain S-11 frontal lobe"
             /dev_stage="34 years old"
             /lab_host="SOLR (kanamycin resistant)"
             /note="Vector: Bluescript SK-; Site_1: EcoRI; Library
             constructed from S-11 frontal lobe, male, 34 years old,
             50% caucasian, 50% Aleutian.   Schizophrenic suicide.
             Random primed into EcoRI site of ZAP II Vector. Mass
             excised. Custom library. Avg insert length 1.4kb.
             Material obtained by Johnston N., Torrey, E.F., Yolken R.,
             and the Stanley Neuropathology Consortium - Analysis of
             RNAs from the Brains of Individuals with Psychiatric
             Diseases (Unpublished) Stanley Neurovirology Laboratory,
             Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT   78 a     99 c    118 g     53 t
ORIGIN

Query Match    9.8%;   Score 333.8; DB 40;  Length 348;
Best Local Similarity 99.1%;  Pred. No. 1.4e-48;
Matches 346;  Conservative  0; Mismatches  2; Indels  1; Gaps  1;

Qy  1259 CTGGGGCGTGTTCCTGCGCCAGTGCGCCATCTCCCTGGTCTTCTCTGGTCATGGACCGGCTG 1318
         ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   348 CTGGGGCGTGTTCCTGCGCCAGTGCGCCATCTCCCTGGTCTTCTCTGGTCATGGACCGGCTG 289

Qy  1319 GTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGGCT 1378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   288 GTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGGCT 229

Qy  1379 GCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGTGCACAGCTTCAGCCGCCCTCACC 1438
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   228 GCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGTGCACAGCTTCAGCCGCCCTCACC 169

Qy  1439 GGGTTCACCTTCTCAGCCCTGCAGATCCTGCCTCCCCTCTACCACCGG 1498
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   168 GGGTTCACCTTCTCAGCCCTGCAGATCCTGCCTCCCCTCTACCACCGG 109

Qy  1499 GAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGAGGTGCTAGCAGTGAGGAC 1558
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   108 GAGAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACACTGAGGTGCTAGCAGTGAGGAC 49

Qy  1559 AGCCTGATGACCAGCTTCCTGCCGCAGGCCCTAAGCCTGGAGCTCCCTTCC 1607
         |||||||||||||||||||||||| |||||||||||||||||||||||||
Db    48 AGCCTGATGACCAGCTTCCTG-CAGGCCCTAAGCCTGGAGCTCCCTTCC 1

RESULT  5
AA225106
LOCUS       AA225106     375 bp      mRNA                      EST       15-AUG-1997
DEFINITION  nc2ldl1.r1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1008789, mRNA
            sequence.
ACCESSION.  .AA225106
```

```
NID          g1846415
VERSION      AA225106.1   GI:1846415
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 375)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Apr 14, 1993 this sequence version replaced gi:692639.

             Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
             M.D., Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html

             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 305.
FEATURES     Location/Qualifiers
     source  1..375
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:1008789"
             /clone_lib="NCI_CGAP_Prl"
             /sex="Male"
             /dev_stage="45 years old"
             /lab_host="DH10B"
             /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
             strand cDNA was primed with oligo(dT)17 on 50 ng of
             DNAse-treated, total cellular RNA obtained from
             5,000-10,000 microdissected, histologically normal
             prostate epithelial cells. Double-stranded cDNA was
             ligated to EcoRI adaptors, 5 cycles of PCR applied to the
             cDNA with an adaptor-specific primer, and the resulting
             PCR product subcloned into pAMP10 by the UDG-cloning
             method (Life Technologies). Average insert size is 600
             bp. NOTE: Not directionally cloned. This library was
             constructed by David Krizman."
BASE COUNT   98 a     94 c     76 g    101 t         6 others
ORIGIN

Query Match     9.5%;     Score 324.2; DB 30; Length 375;
Best Local Similarity   97.0%;   Pred. No. 6.7e-47;
Matches 359;  Conservative   0; Mismatches   8; Indels  3; Gaps  3;

Qy  2966 GGCTGATGAAGGCACTGCCCAAAATTCCCTACCCCAACTTTCCCTACCCCAACTT 3025
         |||||||||| ||||||||||||||| |||| ||||||||||||||||||||||
Db     6 GNCTGATGAAGGCACTGCCCAAAATTTNCCCTANCCCCAACTTTCCCTACCCCAACTT 65

Qy  3026 TCCCCACCAGCTCCACAACCCTGTTGGAGCTACTGCAGGACCACGAAAGCACAAAGTGCGG 3085
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66 TCCCCACCAGCTCCACAACCCTGTTGGAGCTACTGCAGGACCACGAAGCACAAAGTGGCT 125

Qy  3086 TTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACA 3145
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 NTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACA 185

Qy  3146 CAGAAACTCAGGAGCACCCCTGCC-TGAGCTAAGGAGGTGGAGGTCTATCTCTCAGGGGGGT 3204
         |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   186 CAGAAACTCAGGAGCACCCCCTGCCGTGAGCTAAGGAGGTCTATCTCTCAGGGGGGT 245

Qy  3205 TTAAGTGCCGTTGCAATAATGTCGTCTTATTTATTTAGC-GGGTGAATATTTATACT 3263
```

```
Db    246  TTAAGTGCCGTTGCAATAAATGTCCGTCTCTTATTTATTTAGCNGGGGTGAATATTTTATACT  305
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3264  GTAAGTGAGCAATCAGAGTATAAATGTTATGTGACAAAATT-AAAGGCTTCTTATATG  3322
           |||||||||||||||||||||||||||||||||||| ||||||||||||
Db    305  GTAAGTGAGCAATCAGAGTATAAATGTTATGTGACAAAATTAAAAGGCTTTCTTATATG  365
Qy   3323  TTTAAAAAAA  3332
           ||||| |||
Db    366  TTTAANAAAA  375

RESULT  6
AA640153/c
LOCUS       AA640153       315 bp    mRNA          EST    23-OCT-1997
DEFINITION  np28b03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1117613 3',
            mRNA sequence.
ACCESSION   AA640153
NID         g2563932
VERSION     AA640153.1  GI:2563932
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 315)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:430583.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd, ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..315
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="21"
                     /clone="IMAGE:1117613"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT     89 a     69 c     77 g     80 t
ORIGIN

Query Match       9.2%;  Score 314;  DB 36;  Length 315;
Best Local Similarity  100.0%;  Pred. No. 3.7e-45;
Matches  314;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   3011  CCCTACCCCCAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAG  3070
           |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    315  CCCTACCCCCAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAG  256
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3071  AAGCACAAAGTGCGGTTCCCAAGCCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC  3130
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255  AAGCACAAAGTGCGGTTCCCAAGCCTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC  196
Qy   3131  TTGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGAGGTCTTAT  3190
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195  TTGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGAGGTCTTAT  136
Qy   3191  CTCTCAGGGGGGTTAAGTGCCGTTGCAATAATGTCGTCTTATTTATTTAGCGGGGTG  3250
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135  CTCTCAGGGGGGTTAAGTGCCGTTGCAATAATGTCGTCTTATTTATTTAGCGGGGTG  76
Qy   3251  AATATTTTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGTGACAAAATTAAAGG  3310
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75   AATATTTTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGTGACAAAATTAAAGG  16
Qy   3311  CTTTCTTATATGTT  3324
           ||||||||||||||
Db    15   CTTTCTTATATGTT  2

RESULT  7
AI587483/c
LOCUS       AI587483       313 bp    mRNA          EST    14-MAY-1999
DEFINITION  tr51c10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2221842 3',
            mRNA sequence.
ACCESSION   AI587483
NID         g4573924
VERSION     AI587483.1  GI:4573924
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189383.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1960   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 310
            POLYA=NO.
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2221842"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT     87 a     69 c     76 g     81 t
ORIGIN

Query Match       9.1%;  Score 311;  DB 48;  Length 313;
```

RESULT 8 (AA631024/c)

Best Local Similarity 100.0%; Pred. No. 1.2e-44;
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 3012 CCTACCCCAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGA 3071
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313 CCTACCCCAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGA 254

Qy 3072 AGCACAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCT 3131
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253 AGCACAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCT 194

Qy 3132 TGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATC 3191
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193 TGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATC 134

Qy 3192 TCTCAGGGGGGGTTAAGTGCCGTTTGCAATAAATGTCGTCTTATTTATTATTAGCGGGGTGA 3251
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133 TCTCAGGGGGGGTTAAGTGCCGTTTGCAATAAATGTCGTCTTATTTATTATTAGCGGGGTGA 74

Qy 3252 ATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGC 3311
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 ATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGC 14

Qy 3312 TTTCTTATATG 3322
        |||||||||||
Db   13 TTTCTTATATG 3
```

RESULT 8
AA631024/c
```
LOCUS       AA631024          306 bp    mRNA            EST       31-OCT-1997
DEFINITION  nq76g11.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1158308 3',
            mRNA sequence.
ACCESSION   AA631024
NID         g2553635
VERSION     AA631024.1  GI:2553635
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 306)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397799.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image.html

            Insert Length: 375   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..306
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1158308"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
```

---

from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."

```
BASE COUNT     89 a     69 c     70 g     78 t
ORIGIN
```

Query Match   9.0%; Score 306; DB 36; Length 306;
Best Local Similarity 100.0%; Pred. No. 8.9e-44;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 3020 CAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAAA 3079
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  306 CAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAAA 247

Qy 3080 GTGCGGTTCCCAAGCCTTTGCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGGAAT 3139
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  246 GTGCGGTTCCCAAGCCTTTGCCATCTCAGCCCCCAGAGTATATCTGTGTTGGGGAAT 187

Qy 3140 CTCACACAGAAACTCAGGAGGACCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGG 3199
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  186 CTCACACAGAAACTCAGGAGGACCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGG 127

Qy 3200 GGGGTTAAGTGCCGTTTGCAATAAATGCTCTTATTTATTTATTAGCGGGGTGAATATTTTA 3259
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126 GGGGTTAAGTGCCGTTTGCAATAAATGCTCTTATTTATTTATTAGCGGGGTGAATATTTTA 67

Qy 3260 TACTGTAAGTGAGCAATCAGATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTAT 3319
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 TACTGTAAGTGAGCAATCAGATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTAT 7

Qy 3320 ATGTTT 3325
        ||||||
Db    6 ATGTTT 1
```

RESULT 9
AI525162
```
LOCUS       AI525162          674 bp    mRNA            EST       18-MAR-1999
DEFINITION  promrna-9_C08.r bvtumor Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   AI525162
NID         g4439297
VERSION     AI525162.1  GI:4439297
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 674)
AUTHORS     Huang,G.M., Ng,N.W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun
            Yu,J. and Hood,L.
TITLE       Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3136548.

            Contact: Guyang Matthew Huang
            Leroy Hood
            University of Washington
            Department of Molecular Biotechnology, Box 357730, University of
            Washington, Seattle, WA 98195
            Tel: 5106280100
            Fax: 5106280108
            Email: huangm@yahoo.com.
FEATURES             Location/Qualifiers
     source          1..674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="bvtumor"
                     /map="21"
                     /note="Organ: Prostate; Vector: pBluescript; Directional
```

cDNA library was constructed using Lambda zP II kit (Stratagene). mRNA was extracted from a fresh prostate tumor tissues (Urology Department, University of Washington)."

BASE COUNT   125 a   124 c   164 g   178 t   83 others
ORIGIN

Query Match   8.8%;   Score 300.2;   DB 47;   Length 674;
Best Local Similarity   77.2%;   Pred. No.1e-42;
Matches   447;   Conservative   0;   Mismatches 120;   Indels   12;   Gaps   10;

```
Qy   1919  AGGGACTTGGCCAAAATCAGCGTAGAAAACTTCCAGCACATTGGGGTGGAGGGCCTGC  1978
           |||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db     16  AGCGACTTGGCCAAAATACTCAGCGTAGAAAACTTCCAGCACATTGGGGTGGAGGGCCTGC  75

Qy   1979  CTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCATGGGGGCTGGCCGGGCTGGCCCCCA  2038
           ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db     76  CTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCATGGGGGCTNGCCGGGCTGGCCGGCCA  135

Qy   2039  GTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCTGCTGCCACCCTGTGCTGCTGAGGTGCG  2098
           |||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db    136  GTTTCTGTTGCTGCCAAAGTAATGTGGCTNTCTGCTCCTGCCACCNTGTGCTGCTGAGGTGCG  195

Qy   2099  TAGCTGCCACAGCTGGGGGCTGGGGCGTCCCGCTCCCTCCTCCCCCAGTCTCTCAGGGCTGCC  2158
           ||  |||||||||||||||  ||||  ||||||||||||||||||||  ||||||||||||
Db    196  TA--TGCACAGCTGGGGNCTGNGNCGTCCCNTCCCTCCTCCTCNCCCCAGTNTCTCAGGGCTGNC  253

Qy   2159  TGACTGGAGGCCTTCCAAGGGGTTTCAGTCT-GGACTTATACAGGGAGGCCAGAGAGGG-  2217
           |||||||||||||||||||||||||  ||||  |||||||||||||||||||||||||
Db    254  TGACTGGAGGNCTTNCAAGTGGNTTTCAGTCTNTNGGACTTATACAGNAGGCCAGTTGGGT  313

Qy   2217  CTCCATGCACT-GGAATGCGGGGACTCTGCAGGTGGATT-ACCCAGGCTCAGGGTTAAC-  2274
           ||||||||||  ||||||||||||  |||||||||| |  |||||||||||||||  ||
Db    314  TTCCATGCACTGGGAATGCGGGGACTTTGCAGTTGGATTAACCCAGGNTCAGGGTTAANA  373

Qy   2274  AGCTAGCCTCCTAGTTGAGAC-ACACCTAGAGAGAAGGGTTTT-GGGAGCTGAATAAACTC  2331
           |||||||||||||||||||||  |||  ||||||||||||||  ||  |||||||||||||
Db    374  AGTTAGCCTCCTAGTNGNGACAACACACNTAGTTGAAGGGTTTTNGGGNGCTNAATAAACTG  433

Qy   2332  AGTCACCTGGTTTCCCATCTCT--AAGCCCCTAACCTGCAGCTTCGTTTAATGTAGCTC  2389
           |||  ||||||||||||| |||  |||||||||  |||||||||||  |||||||||
Db    434  AGTNACCTGNGTNTCCANTTTNTAAAGNCNTTAAICTGNAAGTNTTNTTNAANTTNAGGT  493

Qy   2390  TTGCATGGGAGTTTCTAGGATGAAACACTCCTCTCCATGGGATTTGAACATATGACTTATTT  2449
           ||||||||||||||||||  |||||  |||||||||||| |  ||||||| |  |||||
Db    494  NTNTNAATNGGGGGTTTTAGTGNAANATNNNNNGGGTTTG-ANANTTNANGTATTT  552

Qy   2450  GTAGGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAG  2488
           |||||||  |||| |  ||  |  ||| ||  |
Db    553  TTAGGNNANGAGCCTTNGGNANAANNNNCCAGGCCNNAG  591
```

RESULT 10
N95796/c
LOCUS   N95796   346 bp   mRNA   EST   20-AUG-1996
DEFINITION   zb66h02.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
     IMAGE:308595 3', mRNA sequence.
ACCESSION   N95796
NID   g1268140
VERSION   N95796.1   GI:1268140
KEYWORDS   EST.
SOURCE   human.
   ORGANISM   Homo sapiens
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
     Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 346)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
     Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
     Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
     Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
     Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
     Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.

and Marra,M.
COMMENT   On May 8, 1995 this sequence version replaced gi:801262.

     Contact: Wilson RK
     Washington University School of Medicine
     4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
     Tel: 314 286 1800
     Fax: 314 286 1810
     Email: est@watson.wustl.edu
     This clone is available royalty-free through LLNL ; contact the
     IMAGE Consortium (info@image.llnl.gov) for further information.
     Insert Length: 622   Std Error: 0.00
     Seq primer: mob.REGA+ET
     High quality sequence stop: 263.
FEATURES   Location/Qualifiers
     source   1..346
                /organism="Homo sapiens"
                /db_xref="GDB:1252008"
                /db_xref="taxon:9606"
                /clone="IMAGE:308595"
                /clone_lib="Soares_fetal_lung_NbHL19W"
                /dev_stage="19 weeks"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer
                [5'-TGTTACCAATCTGAAGTGGGACGGCGCCGCCAATTTTTTTTTTTTTTTTTTT-3'],
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pT7T3 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot - 5. Library constructed by Bento
                Soares and M.Fatima Bonaldo. This library was constructed
                from the same fetus as the fetal heart library, Soares
                fetal heart NbHHl9W."
BASE COUNT   99 a   69 c   87 g   87 t   4 others
ORIGIN

Query Match   8.7%;   Score 295.6;   DB 25;   Length 346;
Best Local Similarity   92.5%;   Pred. No. 5.7e-42;
Matches   320;   Conservative   0;   Mismatches   23;   Indels   3;   Gaps   1;

```
Qy   2983  CCCAAAATTCCCCTACCCCCAACTTCCCCTACCCCCAACTTCCCCACCAGC---TCC  3039
           |||||| ||||||||||  ||||||||||  ||||||| |  ||||||||||||
Db    346  CCAAAATTTCCCNTACCCCNAAACTNNCCCTACCCCANTTTCCCCACCAGGTTCCA  287

Qy   3040  ACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGACACAAAGTGCGGTTCCCAAGCCTTT  3099
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    286  CAACCCCGTTTGGAGCTACTGCAGGACCAGACCACCAAAGTGCGGTTCCCAAGCCTTT  227

Qy   3100  GTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAGAAAACTCAGGAG  3159
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    226  GTCCATCTCAGCCCCCAGAGTATATCTGTTGGGGAATCTCACACAGAAAACTCAGGAG  167

Qy   3160  CACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCAGGGGGGTTAAGTGCCGTTTGC  3219
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    166  CACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCAGGGGGGTTAAGTGCCGTTTGC  107

Qy   3220  AATAAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAAGTGAGCAATCAG  3279
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    106  AATAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAAGTGAGCAATCAG  47

Qy   3280  AGTATAATGTTATGGTGACAAAATTAAAGGCTTTCTTATATATGTT  3325
           |||||||||||||||||||||||||||||||||||||||||||||
Db     46  AGTATAATGTTATGGTGACAAAATTAAAGGCTTTCTTATATATGTT  1
```

RESULT 11
AI69721/c

```
LOCUS       AI696721       301 bp    mRNA              EST       03-JUN-1999
DEFINITION  wc56d05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2322633 3'
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI696721
NID         g4984621
VERSION     AI696721.1  GI:4984621
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 301)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948764.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck. M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 285.
FEATURES             Location/Qualifiers
     source          1..301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2322633"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT    85 a    69 c     65 g     82 t
ORIGIN
```

```
Query Match            8.3%;   Score 282.6; DB 50;   Length 301;
Best Local Similarity 98.6%;   Pred. No. 9.8e-40;
Matches  285;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

Qy   3043 ACCCTGTTTGGAGCTACTGCAGGAAGCACCAGAAGCACAAAGTGCGGTTCCCAAGCCTTTGTC 3102
          |||  ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db    289 AACCCTGTTGGAGCTACTGCAGGAAGCACCAGAAGCACAAAGTGCGGTTCCCAAGCCTTTGTC 230

Qy   3103 CATCTCAGCCCCCAGAGTATATCTGTCTGGGGAATCTCACACAGAGAAACTCAGGAGCAC 3162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 CATCTCAGCCCCCAGAGTATATCTGTCTGGGGAATCTCACACAGAGAAACTCAGGAGCAC 170

Qy   3163 CCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT 3222
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    169 CCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT 110

Qy   3223 AATGTCGTCTTATTTATTTAGCGGGTGAATATTTATACTGTAAGTGAGCAATCAGAGT 3282
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    109 AATGTCGTCTTATTTATTTAGCGGGTGAATATTTATACTGTAAGTGAGCAATCAGAGT 50

Qy   3283 ATAATGTTTATGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAA 3331
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db     49 ATAATGTTTATGTGACAAAATTAAAAGGCTTTCTTATATGTTTAAAAAA 1
```

```
RESULT 12
AI468280/c
LOCUS       AI468280       299 bp    mRNA              EST       30-MAR-1999
DEFINITION  tg57a01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2112840 3'
            similar to contains element PTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AI468280
NID         g4330370
VERSION     AI468280.1  GI:4330370
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 299)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948787.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck. M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 427      Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 286.
FEATURES             Location/Qualifiers
     source          1..299
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2112840"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT    88 a    69 c     66 g     76 t
ORIGIN
```

```
Query Match            8.3%;   Score 281.8; DB 47;   Length 299;
Best Local Similarity 99.0%;   Pred. No. 1.3e-39;
Matches  294;   Conservative   0;   Mismatches   2;   Indels   1;   Gaps   1;

Qy   3029 CCACCAGCTCCACAACCCTGTTTGGAGCTACTGCGAGGACCAGAGAAGCACAAAGTGCGGTTT 3088
          ||  ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    296 CCCACAGCTCCACAACCCTG-TTGGAGCTACTGCGAGGACCAGAGAAGCACAAAGTGCGGTTT 238
```

ORIGIN

```
Query Match                8.2%;     Score 280.6;  DB 28;   Length 342;
Best Local Similarity  95.8%;     Pred No. 2.2e-39;
Matches 320;  Conservative  0;  Mismatches 10;  Indels  4;  Gaps  3;

Qy   1396  GTCCCACAGTGTGGCCGTGGTGACAGCTT--CAGCCGCGCCCTCACCGGGTTCACCTTCTCA  1453
           ||| |||||||||||||||||||||||| ||  || |||||||||||||||||||||||||||
Db    339  GTCAAACAGTGTGGCCGTGGTGACAAGTTCAGCCGCGCCCTCACCGGGTTCACCTTCTCA  280

Qy   1454  GCCCTGCAGATCTGCCCTACACACTGGCCTCCTCTACCACCGGGAGAAGCAGGTGTTC  1513
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279  GCCCTGCAGATCTGCCCTACACACTGGCCTCCTCTACCACCGGGAGAAGCAGGTGTTC  220

Qy   1514  CTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGC  1573
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219  CTGCCCAAATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGC  160

Qy   1574  TTCC-TGCCAGGCCTAAGCCTGGAGCTCCCTTCCTAATGGACACGTGGTGCTGGAGG  1632
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159  TTCCTTGCCAGGCCTAAGCCTGGAGCTCCCTTCCTAATGGACACGTGGTGCTGGAGG  100

Qy   1633  CAGTGGCTGCTCCCACCTCCACCCGGCCTGCGGGGCCTCTGCCTGTGATGTCTCCGT  1692
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     99  CAGTGGCTGCTCCCACCTCCACCCG-NCTCTGCGGGGCCTCTGCCTGTGATGTCTCCGT  41

Qy   1693  ACGTGTGGTGGTGGTGGAGCCCACCGAGGCCAGG  1726
           ||||||||||||||||||||||||||||||||||
Db     40  ACGTGTGGTGGTGGTGGAGCCCACCGAGGCCCGG  7
```

RESULT 14
AA492342/c
LOCUS   AA492342   264 bp   mRNA   EST   19-AUG-1997
DEFINITION ng81d12.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:941207, mRNA
            sequence.
ACCESSION   AA492342
NID   g2221904
VERSION   AA492342.1 GI:2221904
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 264)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1407518.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 427   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 254.
FEATURES   Location/Qualifiers
     source   1..264
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:941207"
                     /clone_lib="NCI_CGAP_Pr6"
```
Qy   3089  CCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAG  3148
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237  CCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACACAG  178

Qy   3149  AAACTCAGGAGGACGACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAA  3208
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177  AAACTCAGGAGGACGACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAA  118

Qy   3209  GTGCCGTTTGCAATAATGTCGTCTTATTATTTAGGCGGGGTGAATATTTTATACTGTAAG  3268
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    117  GTGCCGTTTGCAATAATGTCGTCTTATTATTTAGGCGGGGTGAATATTTTATACTGTAAG  58

Qy   3269  TGAGCAATCAGGTATAAATGTTATGGTGACAAAAATTAAAGGCTTTCTTATATGTTT  3325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57  TGAGCAATCAGGTATAAATGTTATGGTGACAAAAATTAAAGGCTTTCTTATATGTTT  1
```

RESULT 13
AA112573/c
LOCUS   AA112573   342 bp   mRNA   EST   23-DEC-1997
DEFINITION zm28c12.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
            IMAGE:526966 5', mRNA sequence.
ACCESSION   AA112573
NID   g1665120
VERSION   AA112573.1 GI:1665120
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 342)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,E., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE   Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT   On May 9, 1995 this sequence version replaced gi:802278.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 307.
FEATURES   Location/Qualifiers
     source   1..342
                     /organism="Homo sapiens"
                     /db_xref="GDB:3918395"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:526966"
                     /clone_lib="Stratagene pancreas (#937208)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. Pancreatic adenocarcinoma cell line. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'."
BASE COUNT   67 a   100 c   115 g   59 t   1 others
```

```
                /sex="male"
                /tissue_type="prostate"
                /lab_host="DH10B"
                /note="Vector: pAMP10; mRNA made from prostatic
                intraepithelial neoplasia (low-grade), cDNA made by
                oligo-dT priming. Non-directionally cloned.
                Size-selected on agarose gel, average insert size 600 bp.
                Reference: Krizman et al. (1996) Cancer Research
                56:5380-5383."
BASE COUNT     79 a      58 c      50 g      77 t
ORIGIN

Query Match      7.7%;     Score 262.4;   DB 34;   Length 264;
Best Local Similarity   99.6%;   Pred. No. 2.9e-36;
Matches  263;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy  3076  CAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTCTTGGG  3135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   264  CAAAGTGCGGTTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTCTTGGG  205

Qy  3136  GAATCTCACACAGAAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTC  3195
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   204  GAATCTCACACAGAAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTC  145

Qy  3196  AGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATTTAGCGGGGTGAATAT  3255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   144  AGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATTTAGCGGGGTGAATAT  85

Qy  3256  TTTATACTGTAAGTGACGAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTC  3315
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    84  TTTATACTGTAAGTGACCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTC  25

Qy  3316  TTATATGTTTAAAAAAAAAAAAAAA  3339
          ||||||||||||||||||||
Db    24  TTATATGTTTAAAAAAAAAAAA  1

RESULT 15
AA579735
LOCUS        AA579735         288 bp      mRNA                    EST        03-SEP-1997
DEFINITION   nf39g07.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:916188, mRNA
             sequence.
ACCESSION    AA579735
NID          g2357919
VERSION      AA579735.1   GI:2357919
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 288)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      On Nov 29, 1993 this sequence version replaced gi:636080.

             Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
             M.D., Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html

FEATURES     Seq primer: -40m13 fwd. ET from Amersham.
  source     Location/Qualifiers
             1 ..288
```

```
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:916188"
                /clone_lib="NCI_CGAP_Pr2"
                /sex="Male"
                /dev_stage="45 years old"
                /lab_host="DH10B"
                /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                strand cDNA was primed with oligo(dT)17 on 50 ng of
                DNAse-treated, total cellular RNA obtained from
                5,000-10,000 microdissected preneoplastic cells
                histologically-determined to be prostatic intraepithelial
                neoplasia 2 (PIN2) cells.  Double-stranded cDNA was
                ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                cDNA with an adaptor-specific primer, and the resulting
                PCR product subcloned into pAMP10 by the UDG-cloning
                method (Life Technologies).  Average insert size is 600
                bp. NOTE:Not directionally cloned.  This library was
                constructed by David Krizman."
BASE COUNT     81 a      60 c      67 g      80 t
ORIGIN

Query Match      7.5%;     Score 255.6;   DB 35;   Length 288;
Best Local Similarity   96.5%;   Pred. No. 4.5e-35;
Matches  272;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy  3055  GCTACTGCAGGACCAGAGACACAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCC  3114
          |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GCTCGACCAGGACCAGAGACACAAAGTCGGTCCTCCCAAGCCTTTGTCCATCTCAGCCCC  60

Qy  3115  CAGAGTATATCTGTCTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAG  3174
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CAGAGTATATCTGTCTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAG  120

Qy  3175  CTAAGGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTA  3234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CTAAGGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTA  180

Qy  3235  TTTATTTAGCGGGGTGAATATTTTATACTGTAAGTGACGCAATCAGAGTATAATGTTTATG  3294
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   181  TTTATTTAGCGGGGTGAATATTTTATACTGTAAGTGAGC-ATCAGAGTATAATGTTTATG  239

Qy  3295  GTGACAAAATTAAAGGCTTTCCTTATATGTTTAAAAAAAAAA  3336
          |||||||||||||||||||||||||||||||||||||||
Db   240  GTGACAAAATTAAAGGCTTTCCTTATATGTTTAAAAAAAAAAA  281

Search completed: September 28, 1999, 11:29:14
Job time: 1835 sec
```

This Page Blank (uspto)

```
                GenCore version 4.5
        Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   September 25, 1999, 21:34:17 ; Search time 2910.9 Seconds
                                        (without alignments)
                                        3725.602 Million cell updates/sec

Title:           US-09-030-606-110
Perfect score:   3410
Sequence:        1 GGGAACCAGCCTGCACGCGC.........AAAAAATAAAAAAAAAAA 3410

Scoring table:   IDENTITY_NUC

Searched:        679419 seqs, 1590154680 residues

Database :       GenEmbl:*
          1:  gb_bal:*
          2:  gb_ba2:*
          3:  gb_om:*
          4:  gb_ov:*
          5:  gb_pat:*
          6:  gb_ph:*
          7:  gb_pl1:*
          8:  gb_pl2:*
          9:  gb_prl:*
         10:  gb_pr2:*
         11:  gb_pr3:*
         12:  gb_ro:*
         13:  gb_st:*
         14:  gb_sts:*
         15:  gb_sy:*
         16:  gb_un:*
         17:  gb_vi:*
         18:  em_fun:*
         19:  em_htg:*
         20:  em_hum1:*
         21:  em_hum2:*
         22:  em_in:*
         23:  em_om:*
         24:  em_or:*
         25:  em_ov:*
         26:  em_pat:*
         27:  em_ph:*
         28:  em_pl:*
         29:  em_ro:*
         30:  em_sts:*
         31:  em_sy:*
         32:  em_un:*
         33:  em_vi:*
         34:  gb_htg1:*
         35:  gb_htg2:*
         36:  gb_in1:*
         37:  gb_in2:*
         38:  gb_bal:*
         39:  gb_ba2:*
         40:  em_hum3:*
         41:  em_hum4:*
         42:  gb_pr4:*
```

No.		Score				ID	Description
2		101.2	3.0	1252	10	HSSTA	X82434 H.sapiens m
3		95.8	2.8	1544	4	OLANNEXM3	Y11254 O.latipes m
4		94.6	2.7	688	10	I03321	I03321 Sequence 12
5		93.4	2.7	1694	10	HSM800167	AL049382 Homo sapi
6		93.2	2.7	1794	5	E02349	E02349 cDNA sequen
7		92.8	2.7	625	8	AF049925	AF049925 Petunia x
8		92.8	2.7	2556	42	AF125948	AF125948 Homo sapi
9		92.2	2.7	1500	10	HSM800108	AL050024 Homo sapi
10		91.8	2.7	1500	4	XLNAKATP	Y11587 X.laevis mR
11		91.6	2.7	446	37	AF146743	AF146743 Mesobuthu
12		91.4	2.7	2638	7	ATAJ0058	AJ000058 Arabidops
13		91.2	2.7	1887	10	HSGLCNACT	Z82022 H.sapiens m
14		91	2.7	1748	42	AF118274	AF118274 Homo sapi
15	c	90.8	2.7	13414	5	A58523	A58523 Sequence 27
16	c	90.8	2.7	10288	5	A58524	A58524 Sequence 28
17		90.8	2.7	6671	5	AR011880	AR011880 Sequence 1
18		90.8	2.7	6671	5	I50973	I50973 Sequence 1
19		90.8	2.7	6671	5	I69486	I69486 Sequence 1
20		90.8	2.7	1483	10	HSM000068	AL049283 Homo sapi
21	c	90.8	2.7	8414	15	CVU30496	U30496 Cloning vec
22	c	90.8	2.7	8392	15	CVU30497	U30497 Cloning vec
23		90.6	2.7	3905	5	A03736	A03736 H.sapiens m
24		90.6	2.7	3905	5	A31790	A31790 H.sapiens m
25		90.4	2.7	837	10	S78214	S78214 APC-tumor s
26		90.2	2.6	1412	37	AF083228	AF083228 Caenorhab
27		90	2.6	1368	10	HSY16645	Y16645 Homo sapien
28		89.8	2.6	1474	5	A65341	A65341 Sequence 64
29		89.8	2.6	2462	10	HSM800419	AL050116 Homo sapi
30		89.6	2.6	705	10	HSM800237	AL049452 Homo sapi
31		89.6	2.6	1448	12	AF067728	AF067728 Rattus no
32		89.6	2.6	3581	12	RNSTOP	X93495 R.norvegicu
33		89.6	2.6	4670	42	AF104032	AF104032 Homo sapi
34		89.4	2.6	1559	4	OLJ000937	AJ000937 Oryzias 1
35		89.4	2.6	144	5	I89947	I89947 Sequence 26
36		89.4	2.6	5959	10	HSHBRM	X72889 H.sapiens h
37		89.4	2.6	580	10	HSM800144	AL050277 Homo sapi
38		89.4	2.6	3482	10	HSM800550	AL050393 Homo sapi
39		89.4	2.6	3747	10	HSU42766	U42766 Human neuro
40		89.4	2.6	4915	10	HSZPHSAL2	X98834 H.sapiens m
41		89.4	2.6	1591	12	AF087943	AF087943 Rattus no
42		89.2	2.6	2134	4	XLCCTG	X84990 X.laevis Cc
43		89.2	2.6	868	8	AF049923	AF049923 Petunia x
44		89.2	2.6	1008	8	AF117707	AF117707 Lycopersi
45		89.2	2.6	1429	11	AF097996	AF097996 Homo sapi

```
RESULT 1
HSPE54C06/c
LOCUS       HSPE54C06      406 bp    DNA              STS       18-NOV-1998
DEFINITION  H.sapiens flow-sorted chromosome 1 HindIII fragment, SC1pE54C06,
            sequence tagged site.
ACCESSION   AL033941
NID         g3893713
VERSION     AL033941.1  GI:3893713
KEYWORDS    STS; Anonymous marker; single read.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
AUTHORS     Gregory,S., Kettleborough,R., Langford,C., Ross,M.T. and Hunt,S.E.
TITLE       Direct Submission
JOURNAL     Submitted (18-NOV-1998) The Sanger Centre, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk
COMMENT     Vector: pBSIISK+
            Marker stSG33426FS (Primer A : TGTAGGCCAGGATCTGCAG; Primer B :
            CTCGAGCAGTCTATTTGGCC; amplimer size : 141 bp) was mapped to
            chromosome 1 using Radiation Hybrid
            panel Genebridge 4 (GB4).

ALIGNMENTS
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
c	1	261.4	7.7	406	14	HSPE54C06	AL033941 H.sapiens

```
RESULT  2
HSSTA
LOCUS       HSSTA          1252 bp    mRNA            PRI    05-JAN-1995
DEFINITION  H.sapiens mRNA for emerin.
ACCESSION   X82434
NID         g600618
VERSION     X82434.1  GI:600618
KEYWORDS    emerin; STA gene.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1252)
AUTHORS     Bione,S., Maestrini,E., Rivella,S., Mancini,M., Regis,S., Romeo,G.
            and Toniolo,D.
TITLE       Identification of a novel X-linked gene responsible for
            Emery-Dreifuss muscular dystrophy
JOURNAL     Nat. Genet. 8 (4), 323-327 (1994)
MEDLINE     95201795
REFERENCE   2  (bases 1 to 1252)
AUTHORS     Toniolo,D.
TITLE       Direct Submission
JOURNAL     Submitted (01-NOV-1994) D. Toniolo, Istituto di Genetica Biochemica
            ed Evoluzionistica, CNR, Via Abbiategrasso 207, 27100 Pavia, ITALY
FEATURES             Location/Qualifiers
     source          1..1252
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /cell_type="teratocarcinoma"
                     /cell_line="Human teratocarcinoma"
                     /map="Xq28"
     gene            59..823
                     /gene="STA"
     CDS             59..823
                     /gene="STA"
                     /codon_start=1
                     /product="emerin"
                     /protein_id="CAA57817.1"
                     /db_xref="PID:g600619"
                     /db_xref="GI:600619"
                     /db_xref="SWISS-PROT:P50402"
                     /translation="MDNYADLSDTELTTLLRRYNIPHGPVVGSTRRLYEKKIFEYETQ
                     RRRLSPPSSAAASYSFSDLNSTRGDADMYDLPKKEDALLYQSKGYNDDYYEESYFTT
                     RTYGEPESAGPSRAVRQSVTSFPDADAFHHQVHDDDLLSSEEECKDRERPMYGRDSA
                     YQSITHYRPVSASRSSLDLSYYPTSSSTSFMSSSSSSSWLTRRAIRPENRAPGAGLG
                     QDRQVPLWGQLLLFLVFVIVLFFIYHFMQAEEGNPF"
     polyA_signal    1126..1131
BASE COUNT      313 a    344 c    312 g    283 t
ORIGIN

Query Match             3.0%;   Score 101.2;  DB 10;  Length 1252;
Best Local Similarity  87.9%;  Pred. No. 1e-06;
Matches 109;  Conservative  1;  Mismatches 14;  Indels  0;  Gaps  0;

Qy       3287 TGTTTATGTGACAAAATTAAAGGCTTTCTATATGTTTAAAAAAAAAAAAAAAAAA 3346
              ||  ||| || || ||| ||||| |||| ||| ||||| |||||||||||||||
Db       1111 TTTTTGTGGACACACAATAAAAGCCCCGTTTATTCGTTTAAAAAAAAAAAAAAAAAAAAAAAA 1170

Qy       3347 AAAAAAAAAAAAAAAAAAARARAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAA 3406
              |||||||||||||||||||:||||||:||||||||||||||||| ||||||||
Db       1171 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1230

Qy       3407 AAAA 3410
              ||||
Db       1231 AAAA 1234
```

```
RESULT  3
OLANNEXM3
LOCUS       OLANNEXM3      1544 bp    mRNA            VRT    11-MAY-1999
DEFINITION  O.latipes mRNA for annexin max3.
ACCESSION   Y11254
NID         g3288569
VERSION     Y11254.1  GI:3288569
KEYWORDS    annexin max3; calcium-binding protein; phospholipid-binding
            protein.
SOURCE      Japanese medaka.
ORGANISM    Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
            Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
            Atherinomorpha; Cyprinodontiformes; Adrianichthyoidei;
            Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 1544)
AUTHORS     Osterloh,D., Wittbrodt,J. and Gerke,V.
TITLE       Characterization and developmentally regulated expression of four
            annexins in the killifish medaka
JOURNAL     DNA Cell Biol. 17 (10), 835-847 (1998)
MEDLINE     99025617
REFERENCE   2  (bases 1 to 1544)
AUTHORS     Osterloh,D.
TITLE       Direct Submission
JOURNAL     Submitted (10-FEB-1997) D. Osterloh, University Of Muenster, Dept.
            Of Medical Biochemistry, Von-Esmarch-Strasse 56, D-48149 Muenster,
            FRG
FEATURES             Location/Qualifiers
     source          1..1544
                     /organism="Oryzias latipes"
                     /db_xref="taxon:8090"
                     /dev_stage="embryo"
                     /clone="medaka annexin max 3"
     CDS             60..1073
                     /function="calcium- and phospholipid-binding protein"
                     /note="cDNA contains flanking EcoRI and XhoI sites"
                     /codon_start=1
                     /product="annexin max3"
```

```
FEATURES             Location/Qualifiers
     source          1..406
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /sex="female"
                     /dev_stage="adult"
                     /tissue_type="EBV lymphoblastoid cell line"
                     /clone_lib="SC1pE"
                     /clone="SC1pE54C06"
BASE COUNT       97 a    128 c    115 g     66 t
ORIGIN

Query Match             7.7%;   Score 261.4;  DB 14;  Length 406;
Best Local Similarity  96.1%;  Pred. No. 5.4e-31;
Matches 268;  Conservative  0;  Mismatches 11;  Indels  0;  Gaps  0;

Qy       1237 TGAAGGCGTTCGGATGGAGGGCAGCCTGGGGCGTTCCTGCAGTGCGCCATCTCCCTGGTCTT 1296
              || ||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        395 TGCAGGCGGTTCGGATGGAGGGCAGCCTGGGGCGTTCCTGCAGTGCGCCATCTCCGTGGTCTT 336

Qy       1297 CTCTCTGGTCATGGACCCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAG 1356
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 CTATGTGGTCATGGACCCGGCTGGTGCAGCGATTCGGCACTCGAGCAGTCTATTTGGCCAG 276

Qy       1357 TGTGGCAGCTTTCCCTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGT 1416
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        275 TGTGGCAGCATTCCCTGGCTGCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGGT 216

Qy       1417 GACAGCTTCAGCCGCCCCTCACCGGGTCACCTTCTCAGCCCTCGCGAGATCCTGCCCTACAC 1476
              ||||||||||||||||||||||||||||||| || |||||||||||||||||||||||||
Db        215 GACAGCTTCAGCCGCCCCTCACCGGGTCACCGTTTCAGCCCTCGCGAGATCCTGCCCTACAC 156

Qy       1477 ACTGGCCTCCCTCCCTACCACCGGGAGAAGCAGGTGTTCCT 1515
              |||||||||||| ||||||||||||||||||||||| || |
Db        155 ACTGGCCTCCCTATACACCACCGGGAGAAGCAGGTAGTCAT 117
```

```
                    /protein_id="CAA72124.1"
                    /db_xref="PID:e311341"
                    /db_xref="PID:g3288570"
                    /db_xref="GI:3288570"
                    /db_xref="SPTREMBL:O93446"
                    /translation="MSFIQAFLQQTVYLGMPDDSVLRNEGTVTAAPNFSPSGDAAVLD
                    KAIKAKGVDENTIIEILVKRSNEFQRQQIKEAYQQASGKPLESALKSALKGDLEEVLA
                    LLKTPAQYDAQQLKIAMKGLGTDEDTLIEILASRNNRQIMDLKKAYKEDYKKDEEDI
                    RSDTSGDFRAALLALCKAGRTEGISEQLIDSDARALYEAGEGRKGKDCSVFIEILTTR
                    SGPHLRKYFERYSKYSKVDMAKAIDLEMKGDIESCLTAIVKCSGSRAAFFAEKLNLAM
                    KGKGTRKNILTRIMVARSEIDMKLIKEEYKKNYGKTLYKDILDDTKGDYEKILLALCG
                    GDN"
BASE COUNT     547 a    308 c    354 g    335 t
ORIGIN

Query Match               2.8%;     Score 95.8; DB 4;    Length 1544;
Best Local Similarity    91.7%;     Pred. No. 6.7e-06;
Matches 100; Conservative    1; Mismatches    8; Indels    0; Gaps    0;

Qy   3302 AATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3361
          ||||||||||||||||||||| ||  || ||||||||||||||||||||||||||||||
Db   1409 AAATAAAAGCTTTTTCAAATGATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1468

Qy   3362 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAA 3410
          |||||||:|:|||||||||||||||||||||||| |||||||||
Db   1469 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1517

RESULT 4
I03321
LOCUS       I03321           688 bp ss-DNA                PAT       05-MAR-1993
DEFINITION  Sequence 12 from Patent US 4886747.
ACCESSION   I03321
NID         g270715
VERSION     I03321.1  GI:270715
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Derynck,R.M.A. and Goeddel,D.V.
  TITLE     Nucleic acid encoding TGF-beta, and its uses
  JOURNAL   Patent: US 4886747-A 12 12-DEC-1989;
            Genentech, Inc.;
            South San Francisco, CA
FEATURES             Location/Qualifiers
     source          1..688
                     /organism="unknown"
BASE COUNT     240 a    137 c    138 g    173 t
ORIGIN

Query Match               2.8%;     Score 94.6; DB 5;    Length 688;
Best Local Similarity    90.1%;     Pred. No. 1.1e-05;
Matches 100; Conservative    1; Mismatches   10; Indels    0; Gaps    0;

Qy   3300 AAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3359
          || ||||||||| || || || ||  |  ||||||||||||||||||||||||||||||
Db   571  AACAATAAAGGCCTTATTCTCAAGGTGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 630

Qy   3360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
          ||||||||||||||||||||||||||||||| |||||||||||
Db   631  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 681

RESULT 5
HSM800167
LOCUS       HSM800167      1694 bp    mRNA            PRI       21-MAY-1999
DEFINITION  Homo sapiens mRNA; cDNA DKFZp586K1318 (from clone DKFZp586K1318).
ACCESSION   AL049382
NID         g4500170
VERSION     AL049382.1  GI:4500170
```

```
KEYWORDS    .
SOURCE      human.
            Homo sapiens
  ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1694)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1999) MIPS, Am Klopferspitz 18a D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); sequenced by AGOWA within the cDNA
            sequencing consortium of the German Genome Project. This clone is
            available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..1694
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp586K1318"
                     /clone_lib="586 (synonym: hutel). Vector pSport1; host
                     DH10B; sites NotI + SalI/MluI"
                     /dev_stage="adult"
                     /tissue_type="uterus"
     polyA_signal    1578..1583
     polyA_site      1599
BASE COUNT     672 a    238 c    218 g    566 t
ORIGIN

Query Match               2.7%;     Score 93.4; DB 10;    Length 1694;
Best Local Similarity    82.2%;     Pred. No. 1.5e-05;
Matches 106; Conservative    1; Mismatches   22; Indels    0; Gaps    0;

Qy   3282 TATATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAA 3341
          |||| ||||| |||| |||||||||||||||||||| || || || ||||||||||||
Db   1558 TATATTGTTTACAATTAAAAATAAATAATTTACTACTTAATCCTAAAAAAAAAAAAAA 1617

Qy   3342 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA 3401
          ||||||||||||||||||||||||||:||||||||||||||||||| |||
Db   1618 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1677

Qy   3402 AAAAAAAAA 3410
          ||||||||
Db   1678 AAAAAAAAA 1686

RESULT 6
E02349
LOCUS       E02349         1794 bp    RNA                       29-SEP-1997
DEFINITION  cDNA sequence coding for eta peptide.
ACCESSION   E02349
NID         g2170584
VERSION     E02349.1  GI:2170584
KEYWORDS    JP 1990111796-A/1.
SOURCE      Bos taurus.
  ORGANISM  Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Artiodactyla; Ruminantia; Pecora; Bovoidea; Bovidae; Bovinae; Bos.
            Bos taurus
REFERENCE   1  (bases 1 to 1794)
  AUTHORS   Isobe,T., Okuyama,N. and Takahashi,Y.
  TITLE     PROTEIN ACTIVATING AMINO ACID HYDROXYLASE AND GENE THEREOF
  JOURNAL   Patent: JP 1990111796-A 1 24-APR-1990;
            TONEN CORP
COMMENT     OS   (Bovine)
            PN   JP 1990111796-A/1
            PD   24-APR-1990
            PF   21-OCT-1988 JP 1988264097
            PI   ISOBE TOSHIAKI, OKUYAMA NORIO, TAKAHASHI YASUO PC
            C07K15/06,C07K13/00,C07K15/12,C12N15/12,C12P21/02//A61K37/24, PC
            C07K3/02;
            CC   .strandedness: Single;
```

```
CC     topology: Linear;
CC     hypothetical: No;
CC     anti-sense: No;
FH     *source: tissue_type=cerebellum;
FH     Key             Location/Qualifiers
FT     5'UTR           1..165
FT     CDS             166..906
FT                     /product='Eta peptide'
FT     3'UTR           907..>1794
FT     polyA_signal    1671..1676.
                       /organism="Bos taurus"
                       /db_xref="taxon:9913"
BASE COUNT    506 a    454 c    471 g    363 t
ORIGIN

Query Match           2.7%;   Score 93.2;   DB 5;   Length 1794;
Best Local Similarity   69.4%;   Pred. No. 1.6e-05;
Matches 125;   Conservative 1;   Mismatches 54;   Indels  0;   Gaps  0;

Qy  3231 CTTATTTATTTAGCCGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTT 3290
         |||   ||  ||   ||| | |||||| ||| ||||  || ||  ||  ||  || ||
Db  1601 CATGTCAATCAAGATGGGTGATTATGAAATGCCAGACTCCTAAATTAAATGTTTTGGAAT 1660

Qy  3291 TATGGTGACAAAATTAAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAA 3350
         | | ||| |||  ||   ||             | |||||||||||||||||||||||||
Db  1661 TCAGTGGGTAAATAAAAATGCTGCTTGGGATATTAAAAAAAAAAAAAAAAAAAAAAAAAA 1720

Qy  3351 AAAAAAAAAAAAAAAAAAAAARRAAAAAAAAAAAAAAAAAAAAATAAAAAAAAA 3410
         ||||||||||||||::||||||:||||||||||||||||||||| ||||||||||
Db  1721 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1780


RESULT 7
AF049925
LOCUS       AF049925     625 bp   mRNA              PLN       02-JAN-1999
DEFINITION  Petunia x hybrida PGP214 (PGP214) mRNA, partial cds.
ACCESSION   AF049925
NID         g4105787
VERSION     AF049925.1  GI:4105787
KEYWORDS    
SOURCE      Petunia x hybrida.
ORGANISM    Petunia x hybrida
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
REFERENCE   1  (bases 1 to 625)
AUTHORS     Guyon,V., Astwood,J.D. and Taylor,L.P.
TITLE       Isolation and characterization of petunia germinating pollen cDNAs
            expressed in conditionally male fertile pollen
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 625)
AUTHORS     Guyon,V., Astwood,J.D. and Taylor,L.P.
TITLE       Direct Submission
JOURNAL     Submitted (23-FEB-1998) Genetics and Cell Biology, Washington State
            University, Science Hall, Pullman, WA 99164-4234, USA
FEATURES             Location/Qualifiers
     source          1..625
                     /organism="Petunia x hybrida"
                     /db_xref="taxon:4102"
                     /tissue_type="germinating petunia pollen treated with
                     kaempferol"
     gene            1..625
                     /gene="PGP214"
     CDS             <1..313
                     /gene="PGP214"
                     /note="A, R, L rich protein"
                     /codon_start=2
                     /product="PGP214"
                     /protein_id="AAD02551.1"
                     /db_xref="PID:g4105788"
                     /db_xref="GI:4105788"
                     /translation="DWIEHTKARHERAKEQGEEQNGLAPQEARRANHRAQLIARLAAR
                     RGAIVCIFRRDLPWITRYEARGASYSAPRANFDVIFRIGLGLTYYQAFVNPINTLLLS
                     F"
BASE COUNT    216 a    112 c    134 g    163 t
ORIGIN

Query Match           2.7%;   Score 92.8;   DB 8;   Length 625;
Best Local Similarity   91.5%;   Pred. No. 2e-05;
Matches 97;   Conservative 1;   Mismatches  8;   Indels  0;   Gaps  0;

Qy  3305 TAAAGGCTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3364
         ||| ||| || ||||||                |||||||||||||||||||||||||||
Db   519 TAATGGGTTACTTATTCTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 578

Qy  3365 AAARARAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAA 3410
         |||::||||||||||||||||||||||||| ||||||||||
Db   579 AAARAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 624


RESULT 8
AF125948
LOCUS       AF125948     2556 bp   mRNA              PRI       12-APR-1999
DEFINITION  Homo sapiens DNA repair protein RAD52 delta isoform (RAD52) mRNA,
            alternatively spliced, complete cds.
ACCESSION   AF125948
NID         g4581005
VERSION     AF125948.1  GI:4581005
KEYWORDS    
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2556)
AUTHORS     Kito.K., Wada.H., Yeh,E.T.H. and Kamitani,T.
TITLE       Identification of Novel Isoforms of Human RAD52
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 2556)
AUTHORS     Kito.K., Wada.H., Yeh,E.T.H. and Kamitani,T.
TITLE       Direct Submission
JOURNAL     Submitted (04-FEB-1999) Internal Medicine, University of Texas,
            6431 Fannin, MSB4200, Houston, TX 77030, USA
FEATURES             Location/Qualifiers
     source          1..2556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     gene            1..2556
                     /gene="RAD52"
     CDS             44..400
                     /gene="RAD52"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="DNA repair protein RAD52 delta isoform"
                     /protein_id="AAD24575.1"
                     /db_xref="PID:g4581006"
                     /db_xref="GI:4581006"
                     /translation="MSGTEAILGGRDSHPAAGGGSVLCFGQCQYTAEEYQAIQKALR
                     QRLGPEYISSRMAGGGQKVCYIEGHRVINLANEMFGYNGWAHSITQQNVGEYALQQWG
                     LLHCPAPAESLLWYRR"
BASE COUNT    749 a    581 c    638 g    588 t
ORIGIN

Query Match           2.7%;   Score 92.8;   DB 42;   Length 2556;
Best Local Similarity   76.7%;   Pred. No. 1.8e-05;
Matches 112;   Conservative 1;   Mismatches 33;   Indels  0;   Gaps  0;

Qy  3265 TAAGTGAGCAATCAGATGATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTT 3324
         |  || ||  || | || |  || |||  |||   || |    || ||  |||||||||
Db  2403 TTATTAACAAAATAGAAGAATATATATTGCTTATTGCTTATCCTGAAATGTCTTGAAAA 2462
```

```
Qy   3325  TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3384
           |||||||||||||||||||||||||||||||||||::|||||||||
Db   2463  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2522

Qy   3385  AAAAAAAAAAAATAAAAAAAAAAAAAAAAA  3410
           ||||||||||||||||||||||||
Db   2523  AAAAAAAAAAAAAAAAAAAAAAAAAA  2548

RESULT  9
HSM800108
LOCUS       HSM800108     1500 bp    mRNA               PRI    21-MAY-1999
DEFINITION  Homo sapiens mRNA; cDNA DKFZp564D206 (from clone DKFZp564D206).
ACCESSION   AL050024
NID         g4884093
VERSION     AL050024.1   GI:4884093
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1500)
AUTHORS     Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAY-1999) MIPS, Am Klopferspitz 18a D-82152
            Martinsried, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); sequenced by AGOWA within the cDNA
            sequencing consortium of the German Genome Project. This clone is
            available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..1500
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp564D206"
                     /clone_lib="564 (synonym: hfbr2). Vector pAMP1; host
                     XI-2blue; sites NotI + SalI"
                     /dev_stage="fetal"
                     /tissue_type="brain"
     gene            1..405
                     /gene="DKFZp564D206"
     CDS             <1..405
                     /gene="DKFZp564D206"
                     /note="unknown"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB43243.1"
                     /db_xref="PID:e1454921"
                     /db_xref="PID:g4884094"
                     /db_xref="GI:4884094"
                     /translation="GGCRELRSTKYISDGQCTSISPLKELVCAGECLLLPVLPNWIGG
                     GYGTKYWSRASSQEWRCVNDKTRTQRIQLCQDGSRTYKITVVTACKCKRYTRQHNE
                     SSHNFEMSMSPAKPVCHHRERKRASKSSKHSMS"
     polyA_signal    1390..1395
     polyA_site      1409
BASE COUNT    516 a    288 c    269 g    427 t
ORIGIN

Query Match             2.7%;  Score 92.2;  DB 10;  Length 1500;
Best Local Similarity  83.7%;  Pred. No. 2.4e-05;
Matches  103;  Conservative   1;  Mismatches  19;  Indels  0;  Gaps  0;

Qy   3288  GTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAA  3347
           |||| |||| |  | ||||||||||||||| ||| |||||||||||||||||||||||||
Db   1375  GTTTCTTTAACAAAATAAACACTGCTAAAAGTTAAAAAAAAAAAAAAAAAAAAAAAAAAA  1434

Qy   3348  AAAAAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAATAAA  3407
           ||||||||||||||||||||||::||||||||||||||||||||||||||
Db   1435  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1494
```

```
Qy   3408  AAA  3410
           ||
Db   1495  AAA  1497

RESULT  10
XLNAKATP
LOCUS       XLNAKATP      1500 bp    mRNA               VRT    15-AUG-1997
DEFINITION  X.laevis mRNA for Na,K-ATPase gamma subunit.
ACCESSION   Y11587
NID         g2052283
VERSION     Y11587.1   GI:2052283
KEYWORDS    ATPase; gamma subunit; gamma subunit of sodium potassium ATPase;
            NA+, K+-ATPase.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Vertebrata; Amphibia; Batrachia;
            Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1500)
AUTHORS     Beguin,P., Wang,X., Firsov,D., Puoti,A., Claeys,D.,
            Horisberger,J.D. and Geering,K.
TITLE       The gamma subunit is a specific component of the Na,K-ATPase and
            modulates its transport function
JOURNAL     EMBO J. 16 (14), 4250-4260 (1997)
MEDLINE     97392454
REFERENCE   2  (bases 1 to 1500)
AUTHORS     Beguin,P.
TITLE       Direct Submission
JOURNAL     Submitted (26-FEB-1997) P. Beguin, University of Lausanne,
            Institute of Pharmacology and Toxicology, Rue du Bugnon 27, 1005
            Lausanne, Switzerland
FEATURES             Location/Qualifiers
     source          1..1500
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /cell_line="A6"
                     /tissue_type="kidney"
     CDS             63..248
                     /codon_start=1
                     /product="gamma subunit of Na,K-ATPase"
                     /protein_id="CAA72326.1"
                     /db_xref="PID:e307198"
                     /db_xref="PID:g2052284"
                     /db_xref="GI:2052284"
                     /db_xref="SWISS-PROT:O13001"
                     /translation="MADAQDDMSQMQDKFTVDYETIRKGGLIFAAIAFVVGMLIIFSG
                     RFRCGRKKQLRALNDDM"
BASE COUNT    559 a    249 c    260 g    432 t
ORIGIN

Query Match             2.7%;  Score 91.8;  DB 4;  Length 1500;
Best Local Similarity  81.4%;  Pred. No. 2.7e-05;
Matches  105;  Conservative   1;  Mismatches  23;  Indels  0;  Gaps  0;

Qy   3282  TATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAA  3341
           |||||| ||| ||| |  | |||||||||| ||| ||| ||||||||||||||||||||||
Db   1312  TATACTGTGTATATTGAATTATTAAAGGAGGCTGGTCTCAAAAAAAAAAAAAAAAAAAAA  1371

Qy   3342  AAAAAAAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAATAA  3401
           ||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db   1372  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  1431

Qy   3402  AAAAAAAA  3410
           ||||||||
Db   1432  AAAAAAAA  1440

RESULT  11
AF146743
LOCUS       AF146743      446 bp    mRNA               INV    17-JUN-1999
DEFINITION  Mesobuthus martensii BmK2 mRNA, complete cds.
```

```
ACCESSION   AF146743
NID         g5081730
VERSION     AF146743.1   GI:5081730
KEYWORDS    .
SOURCE      Manchurian scorpion.
  ORGANISM  Mesobuthus martensii
            Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
            Buthoidea; Buthidae; Mesobuthus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Wenxin,L. and Shunyi,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1999) Virology Dept., Wuhan University, Luojia
            Mountain, Wuhan, Hubei Province, People's Republic of China
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Mesobuthus martensii"
                     /db_xref="taxon:34649"
                     /tissue_type="venom gland"
     CDS             7..180
                     /codon_start=1
                     /product="BmK2"
                     /protein_id="AAD39511.1"
                     /db_xref="PID:g5081731"
                     /db_xref="GI:5081731"
                     /translation="MKPRVFFLLFLLVAAMIETGESEENEEGSNESGKSTEAKNTDAS
                     VDNEDSDIDGDSD"
BASE COUNT    260 a     35 c     59 g     92 t
ORIGIN

Query Match             2.7%;  Score 91.6;  DB 37;  Length 446;
Best Local Similarity   80.3%;  Pred. No. 3.2e-05;
Matches  106;  Conservative   1;  Mismatches  25;  Indels   0;  Gaps   0;

Qy   3279  GAGTATAATGTTTATGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAA 3338
                   ||| || |||||||||||  |  | || |||| || |||||||||||||||
Db    239  GAAAAAATCTTTTACTGAAATGAAGTGAATAAAATAATTAATTATTTGCATAAAAAAAAAAAA 298

Qy   3339  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 3398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 358

Qy   3399  AAAAAAAAAA 3410
           ||||||||||
Db    359  AAAAAAAAAA 370

RESULT  12
ATAJ0058
LOCUS       ATAJ0058      2638 bp    mRNA            PLN       06-APR-1998
DEFINITION  Arabidopsis thaliana mRNA for MCM3 homolog.
ACCESSION   AJ000058
NID         g3036818
VERSION     AJ000058.1   GI:3036818
KEYWORDS    MCM3 homolog.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2638)
  AUTHORS   Dodeman,V.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Dodeman V.L., CNRS/ERS 569, Institut de
            Biotechnologie des Plantes, Universite de Paris-Sud Bat 630, 91405
            Orsay CEDEX, FRANCE
REFERENCE   2  (bases 1 to 2638)
  AUTHORS   Dodeman,V.L., Phan,T., Sabelli,P. and Bergounioux,C.
  TITLE     Expression analysis of Arabidopsis thaliana MCM3 homolog during the
            cell cycle
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..2638
```

```
                     /organism="Arabidopsis thaliana"
                     /strain="C0"
                     /sub_species="Columbia"
                     /db_xref="taxon:3702"
                     /tissue_type="seedling"
     CDS             53..2383
                     /codon_start=1
                     /product="MCM3 homolog"
                     /protein_id="CAA03887.1"
                     /db_xref="PID:e1286040"
                     /db_xref="PID:g3036819"
                     /db_xref="GI:3036819"
                     /db_xref="SPTREMBL:O65400"
                     /translation="MDVPEETRLRHKRDFIQFLDSMYMEEIKALVHQKRHRLLIINISD
                     IHHHFREVASRILKNPNEYMQSFCDAATEATRAIDPKYLKEGELVLVGFEGYFVSRVV
                     TPRELLSDFIGSMVCVEGIVTKCSLVRPKVVSVHFCPSTGEFTNRDYRDTTSHAGLP
                     TGSVYPTRDDKGNLLVTEYGLCKVKDHQTLSIQEVPENAAPGQLPRSVDVIAEDDLVD
                     SCKPGDRVSVFGIYKALPGKSKGSVNGVFRTILIANNIALLNKEANAPIYTKQDLDNI
                     KNIARRDDAFDLIARSLAPSIYGHAWIKKAVVLMLGGVEKNLKNGTHLRGDINMMMV
                     GDPSVAKSOLLRAIMNIAPLAISTTGRGSSGVGLTAAVTSDQETGERRLEAGAMVLAD
                     KGIVCIDEFDKMNDQDRVAIHEVMEQQTVTIAKAGIHASLNARCSVVAAANPIYGTYD
                     RSLIPTKNIGLPDSLLSRFDLLFIVLDQMDAGIDSMISEHVLRMHRYKNDRSEAGPDG
                     SLPYAREDNAESEMFVKYNQTLHGKKKRGQTHDKTLTIKFLKKYIHYAKHRITPKLTD
                     EASERIAEZAYADLRNAGSDTKTGGTLPITARTLETIIRLAATHAHAKMKLSSEVTKADAE
                     AALKLMNFAIYHQELTEMDDREQEORQEAQADEQEERTPSGRRGNQRRNNEDGAENDTAN
                     VDSETADPMEVDEPSVEQFSGTVSAARIETEFRVFGQHMRTHRLDDISIADIETVVNN
                     NGVGASRYSADEIMALLEKLQDDNKVMISDGKVHII"
BASE COUNT    887 a     486 c     586 g     679 t
ORIGIN

Query Match             2.7%;  Score 91.4;  DB 7;  Length 2638;
Best Local Similarity   73.0%;  Pred. No. 3e-05;
Matches  116;  Conservative   1;  Mismatches  42;  Indels   0;  Gaps   0;

Qy   3252  ATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGC 3311
           || || |||   || || |||  |     |||  || || || || || |||| |||
Db   2475  ATTTTGTAAAATCTTTATTCATTACGTTATGTTTTTTGCATTTACTAGTGAAGTAAAACTG 2534

Qy   3312  TTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAARAA 3371
               | |    |||| ||  |||||||||||||||||||||||||||||||||||||:||
Db   2535  ATGATAAAATGCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2594

Qy   3372  AAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAATAAAAAAAAAAA 3410
           ||||||||||||||||||||||||||  |||||||||||||||||||
Db   2595  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2633

RESULT  13
HSGLCNACT
LOCUS       HSGLCNACT     1887 bp    mRNA            PRI       01-JUL-1997
DEFINITION  H.sapiens mRNA for GlcNac-1-P transferase.
ACCESSION   Z82022
NID         g2239118
VERSION     Z82022.1   GI:2239118
KEYWORDS    GlcNac-1-P transferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1887)
  AUTHORS   Eckert,V., Mazhari-Tabrizi,R., Blank,M., Mumberg,D., Funk,M. and
            Schwarz,R.
  TITLE     Cloning and functional expression of the human GlcNac-1-P
            transferase, the enzyme for the committed step of the
            dolichol-cycle by heterologous complementation in yeast
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1887)
  AUTHORS   Eckert,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1996) Eckert V., University of Marburg,
            Medizinisches Zentrum fuer Hygiene, Robert-Koch-Str. 17, Marburg,
            Germany, D-35037
```

FEATURES　　　　　Location/Qualifiers
　source　　　　　1..1887
　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　　/clone="hsalg7"
　　　　　　　　　/tissue_type="lung"
　　　　　　　　　/cell_type="fibroblast"
　　　　　　　　　/clone_lib="cDNA library (pRS416-vector)"
　CDS　　　　　　 104..1306
　　　　　　　　　/codon_start=1
　　　　　　　　　/product="GlcNac-1-p transferase"
　　　　　　　　　/protein_id="CAB04787.1"
　　　　　　　　　/db_xref="PID:e280927"
　　　　　　　　　/db_xref="PID:g2239919"
　　　　　　　　　/db_xref="GI:2239919"
　　　　　　　　　/db_xref="SPTREMBL:O15216"
　　　　　　　　　/translation="MPLLNLIVSLLGFVATVTLIPAFLGHFIAARLCGQDLNKTSRQ
　　　　　　　　　QIPESQGVISGAVFLIILFCFIPPFLNCFVKEQCKAFPHHEFVALIGALLAICCMIF
　　　　　　　　　LGFADDVLNLRWRHKLLLPTAASLPLLMVYFTNFGNTTIVPKFRPILGLHLDIGIL
　　　　　　　　　YYVMGLLVLIVFCTNAINILAGINGLEAGQSLVISASIIVFNLVELEGDCDDHVFSLY
　　　　　　　　　FMIPFFFTTLGLLYHNWPSRVFVGDTFCYFAGMTFAVVGILGHFSKTMLLFMPQVF
　　　　　　　　　NFIYSLPQLLHIIPCPRHRIPRLNIKTGKLEMSYSKFKTKSLSFPLGTFILKVAESLQL
　　　　　　　　　VTVHQSETEDGEFTECNNMTLINLLKVLGPIHERNLTLLLLLQILGSAITFSIRYQ
　　　　　　　　　LVRLFYDV"
BASE COUNT　　　465 a　　　514 c　　　372 g　　　536 t
ORIGIN

Query Match
Best Local Similarity　　2.7%;　　　Score 91.2;　　DB 10;　　Length 1887;
　　Matches 99;　　Conservative　　1;　　Mismatches　　14;　　Indels　　0;　　Gaps　　0;

Qy　3297　GACAAAATAAAGGCTTTCTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3356
　　　　　　 || ||| |||| ||| |||| ||||||||||||||||||||||||||||||||||
Db　1767　GAGACAATAAGTGTTGTGTCTTTTTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1826

Qy　3357　AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
　　　　　　 ||||||||||||||||||||||||||||||||| ||||||||||
Db　1827　AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1880

RESULT 14
AF118274
LOCUS　　　　　 AF118274　　　　1748 bp　　　mRNA　　　　　　　 PRI　　　25-APR-1999
DEFINITION　　 Homo sapiens DNb-5 mRNA, partial cds.
ACCESSION　　　AF118274
NID　　　　　　 g4680228
VERSION　　　　 AF118274.1　　GI:4680228
KEYWORDS　　　 .
SOURCE　　　　　human.
ORGANISM　　　 Homo sapiens
　　　　　　　　 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
　　　　　　　　 Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　　　1　(bases 1 to 1748)
AUTHORS　　　　 Amler,L.C., Bauer,A., Corvi,R., Dihlmann,S., Praml,C.,
　　　　　　　　 Cavenee,W.K., Schwab,M. and Hampton,G.M.
TITLE　　　　　 Identification and Characterization of Novel Genes Located at the
　　　　　　　　 t(1;15)(p36.2;q24) Translocation Breakpoint in the Neuroblastoma
　　　　　　　　 Cell Line NGP
JOURNAL　　　　 Unpublished
REFERENCE　　　2　(bases 1 to 1748)
AUTHORS　　　　 Amler,L.C. and Hampton,G.M.
TITLE　　　　　 Direct Submission
JOURNAL　　　　 Submitted (06-JAN-1999) Genos Biosciences, 11099 North Torrey Pines
　　　　　　　　 Road, La Jolla, CA 92037, USA
FEATURES　　　 Location/Qualifiers
　source　　　　1..1748
　　　　　　　　 /organism="Homo sapiens"
　　　　　　　　 /db_xref="taxon:9606"
　　　　　　　　 /chromosome="1"
　　　　　　　　 /map="1p36.1-p36.2"
　　　　　　　　 /note="located close to the reciprocal translocation
　　　　　　　　 (1;15)(p36.2;q24) in neuroblastoma cell line NGP"

CDS　　　　　　　<1..1602
　　　　　　　　 /note="deleted in neuroblastoma."
　　　　　　　　 /codon_start=1
　　　　　　　　 /product="DNb-5"
　　　　　　　　 /protein_id="AAD27583.1"
　　　　　　　　 /db_xref="PID:g4680229"
　　　　　　　　 /db_xref="GI:4680229"
　　　　　　　　 /translation="KNGPTAFGVLLPXXGTQTLSLSPXFGGGFGYVLGGIHWDKTGFG
　　　　　　　　 RALGQOFRVIXLFTAVTLSVTTVLTLVSIPERPLRPPSEKRAAMKSPSLPLPPSPVL
　　　　　　　　 PEEGPGDSLFSHTANFSSFISPPSPLTPKYGSFISRDSSLTGISEFASSFGTANIDS
　　　　　　　　 VLIDCFTGHDSYLAIPGSVPRPPIVSFPRAPDGFYRQDRGLLEGREGALTSGCDGD
　　　　　　　　 ILRVGSLDTSKPRSSGILKRPQTLAIPDAAGGGPETSRRNVTFSQQVANILLNGVK
　　　　　　　　 YESELTGSSERAEQPLSVGRLCSTICNMPKALRTLCVNHFLGWLSFEGMLLFYTDFMG
　　　　　　　　 EVVFQQDPKAPHTSEAYQKYNSGVTMGCWGMCIYAFSAAFYSAILEKLEEFLSVRTLY
　　　　　　　　 FIAYLAFGLGTGLAATLSRNLYVVLSLCITYGILFSTLCTLPYSLLCDYYQSKKFAGSS
　　　　　　　　 ADGTRRGMGVDISLLSCQYFLAQILVSLVLGPLTSAVGSANGVMYFSSLVSFLGCLYS
　　　　　　　　 SLFVIYEIPPSDAADEHRPLLLNV"
BASE COUNT　　 318 a　　　578 c　　　504 g　　　341 t　　　　　7 others
ORIGIN

Query Match
Best Local Similarity　　50.6%;　　Score 91;　　DB 42;　　Length 1748;
　Matches 220;　　Conservative　　0;　　Mismatches　　215;　　Indels　　0;　　Gaps　　0;

Qy　1075　CTGCCGCATGCCCGGCACCCTGCGCCGCTCTTCGTGCTGCTGAGCTGCTGCAGCTGGATGGC 1134
　　　　　　 ||||| ||||||| || || ||||||||||||||||||||||||| |||||||||| |||
Db　903　CTGCAACATGCCCAAGGCGCTACGCCACCCACTTCCTGGGGTGGCTCTC 962

Qy　1135　ACTCATGACCTTCACGCTGTTTACCAGGGATTCGTGGGCGGAGGGGCGTGTACCCAGGGCGT 1194
　　　　　　 | ||| ||||||||||||| || ||| |||||| ||||||| || |||||
Db　963　ATTCGAGGGGATGTTGCCTTCTACACGAGCTTCATGGGCGAGGTGGTTTCAGGGGGA 1022

Qy　1195　GCCCAGAGTCGAGCCGGGCACCGAGGCCCGGAGAGCACTATGATGAAGGCGGTTCGGATGGG 1254
　　　　　　 |||| ||| ||| || || | |||||| || ||| || ||||| |||||||
Db　1023　CCCCAAGGCCCCGCACACATCAGCGGTATCAGAGTACACACAGGGCGTGACCATGGG 1082

Qy　1255　CAGCCTGGGGCTGTTCCTGCAGTGCCGCCATCTCCCTGGCTCTTCTCGTGCTCATGGGACCG 1314
　　　　　　 || ||| || | ||| || || || |||| | | |||||||| ||| |||||||
Db　1083　CTGCTCGGGGCATGTGATCTACGCCTTCAGTGCTGCCTTCTACTCAGCTATCCTGGAGAA 1142

Qy　1315　GCTGGTGCAGGGATTCGGCACTCGAGCAGTCTATTGGCCAGTGTGGCAGCTTCCCTGT 1374
　　　　　　 |||||| ||| || ||| |||| | ||||| | |||||||||| ||| |||
Db　1143　GCTGGAGAGTTCCTCAGCGTCCGCACCCTCACTTCACGCCGTCATCGCCCTTCGGCCT 1202

Qy　1375　GGCTGCCGGTGCGCACATGCCGTCCCACAGTGTGGCCGTGGTGACAGCTTCAGCCGCCCT 1434
　　　　　　 || ||| ||| ||||| |||| | ||| | ||||| | |||||
Db　1203　GGGACCGGGCTTGCCACCCTCCAGGAACCTCACGTGGTCCTCGCGTCTGCATAAC 1262

Qy　1435　CACCGGGTTCACCTTCACGCTGGGCCTCGCGAGATCCGCACACTGCCTGCCTCCTCCACCA 1494
　　　　　　 | |||| ||||||| |||| |||||||| ||| |||| ||| || |||
Db　1263　CTACGGGATTTTATTTTCCACCCTGTGCACCTTGCCTGCTCTCGCTGCTCTGCGATTACTA 1322

Qy　1495　CCGGGAGAAGCCAGGT 1509
　　　　　　 || ||| |||||||
Db　1323　TCAGAGTAAGAAGTT 1337

RESULT 15
A58523/c
LOCUS　　　　　 A58523　　　　13414 bp　　　DNA　　　circular　　PAT　　05-MAR-1998
DEFINITION　　 Sequence 27 from Patent WO9638555.
ACCESSION　　　A58523
NID　　　　　　 g3714139
VERSION　　　　 A58523.1　　GI:3714139
KEYWORDS　　　 .
SOURCE　　　　　unidentified.
ORGANISM　　　 unidentified
　　　　　　　　 unclassified.
REFERENCE　　　1　(bases 1 to 13414)
AUTHORS　　　　 Bogaert,T., Stringham,E. and Vandekerckhove,J.
TITLE　　　　　 PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH CONTROL CELL
　　　　　　　　 BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND PHARMACEUTICAL COMPOSITIONS

```
JOURNAL     CONTAINING THEM AND THEIR USE IN THE CONTROL OF CELL BEHAVIOUR
            Patent: WO 9638555-A 27 05-DEC-1996;
            BOGAERT THIERRY (BE)
COMMENT     Other publication AU 6123496 961218.
FEATURES             Location/Qualifiers
     source          1..13414
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT   3854 a   3072 c   2884 g   3603 t         1 others
ORIGIN

Query Match
Best Local Similarity   2.7%;   Score 90.8; DB 5;  Length 13414;
Matches  98; Conservative   1; Mismatches  13;  Indels    0;  Gaps   0;

Qy   3299 CAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3358
          ||||| || ||| ||| || | |||||||||||||||||||||||||||||||||
Db  11234 CAACATCTACTTCTGTATTATTTCTATGAAAAAAAAAAAAAAAAAAAAAAAAAAAA 11175

Qy   3359 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAA 3410
          ||||||||:|||||||||||||||||||||| |||||||||||
Db  11174 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 11123


Search completed: September 25, 1999, 23:23:04
Job time: 6527 sec
```

OM nucleic - nucleic search, using sw model

Copyright   GenCore version 4.5
Copyright (c) 1993 - 1998   Compugen Ltd.

Run on: September 25, 1999, 07:42:29 ; Search time 273 Seconds
                              (without alignments)
                              3125.110 Million cell updates/sec

Title: US-09-030-606-110
Perfect score: 3410
Sequence: 1 GGGAACCAGCCTGCCACGCGC.........AAAAAAATAAAAAAAAAAA 3410

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database :   N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3409.6	100.0	3410	1	V58586	Prostate tumour sp
2	3409.6	100.0	3410	1	V61201	Full length cDNA s
3	2136.4	62.7	2152	1	V71181	Consensus sequence
4	2114.8	62.0	2143	1	V71180	Clone 1711346IH, t
5	673.4	19.7	789	1	V58487	3' fragment of pro
6	673.4	19.7	789	1	V61144	3' cDNA sequence o
c 7	603	17.7	772	1	V58488	5' fragment of pro
c 8	603	17.7	772	1	V61145	5' cDNA sequence o
9	405.8	11.9	435	1	X40506	Human secreted pro
10	319.8	9.4	342	1	V71173	PS108 gene-specifi
11	291	8.5	294	1	V71177	PS108 gene-specifi
12	287	8.4	288	1	V71175	PS108 gene-specifi
13	270	7.9	272	1	V71176	PS108 gene-specifi
14	265	7.8	265	1	V71174	PS108 gene-specifi
15	256.4	7.5	258	1	V71166	PS108 gene-specifi
16	254	7.4	255	1	V71168	PS108 gene-specifi
17	247	7.2	247	1	V71169	PS108 gene-specifi
18	220	6.5	231	1	V71170	PS108 gene-specifi
19	217	6.4	217	1	V71167	PS108 gene-specifi
20	213	6.2	213	1	V71179	PS108 gene-specifi
21	210	6.2	223	1	V71172	PS108 gene-specifi
22	183	5.4	195	1	V71171	PS108 gene-specifi
23	149	4.4	151	1	V71178	PS108 gene-specifi
24	97.4	2.9	1582	1	T18831	Human survival mot
25	95.2	2.9	1582	1	T28259	Survival motor neu
26	95.2	2.8	1376	1	X04325	Human secreted pro
27	94.6	2.8	2676	1	Q02819	cDNA sequence enco
28	94.6	2.8	2671	1	Q03303	Entire porcine tra
29	94.6	2.8	2669	1	Q56925	Pig TGF-beta-3. Nu
30	94.4	2.8	2447	1	V54587	Human secretory pr
31	94.4	2.8	3899	1	V63192	cDNA from clone dt
32	93.2	2.7	1834	1	Q04690	Encodes Mammalian
33	93	2.7	2427	1	Q04107	Human pro-urokinas
34	92.6	2.7	2123	1	T59701	PTH-like peptide D
35	92.2	2.7	6020	1	Q06648	Plasminogen gene f
36	92.2	2.7	6010	1	Q11998	Human plasminogen
37	91.6	2.7	921	1	V40521	Homo sapiens CH27_
38	91.4	2.7	1066	1	Q49943	Human anti-HBs lig
39	91.4	2.7	7753	1	Q56826	Norwalk virus stra
40	90.8	2.7	6671	1	Q97228	Human DRADA gene 6
c 41	90.8	2.7	13414	1	T71321	Plasmid pCB50 enco
c 42	90.8	2.7	10288	1	T71322	Plasmid pCB51 enco
43	90.8	2.7	6671	1	T89799	Human double stran

| 44 | 90.8 | 2.7 | 6671 | 1 | V27063 | Human double-stran |
| 45 | 90.8 | 2.7 | 3213 | 1 | V55742 | Human secreted pro |

ALIGNMENTS

RESULT 1
V58586
ID V58586 standard; cDNA; 3410 BP.
AC V58586;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone L1-12.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
FH Key   Location/Qualifiers
FT CDS   284..1945
FT   /*tag= a
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69385.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 84-85; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 3410 BP; 667 A; 1014 C; 945 G; 783 T;

Query Match   100.0%; Score 3409.6; DB 1; Length 3410;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 3410; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 GGGAACCAGCCTGCACGCGCTGGCTCCGGTGACAGCCGCGCCTCGGCCAGGATCTGA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGAACCAGCCTGCACGCGCTGGCTCCGGTGACAGCCGCGCCTCGGCCAGGATCTGA 60

Qy   61 GTGATGAGACGTGTCCCACTGAGGTGCCCACAGCAGGTGTGAGCATGGGCTGAG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTGATGAGACGTGTCCCACTGAGGTGCCCACAGCAGGTGTGAGCATGGGCTGAG 120

Qy   121 AAGCTGGACCGGCACCAAAGGGCTGGCAGAAATGGGCGCCTGGCTGATTCCTAGGCAGTT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAGCTGGACCGGCACCAAAGGGCTGGCAGAAATGGGCGCCTGGCTGATTCCTAGGCAGTT 180

Qy   181 GGCGGCAGCAAGGAGGAGAGGCCCAGCTTCTGGAGCAAGCCGAGACGAAGCAGTTCTG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCGGCAGCAAGGAGGAGAGGCCCAGCTTCTGGAGCAAGCCGAGACGAAGCAGTTCTG 240

Qy   241 GAGTGCCTGAACGGCCCCTGAGCCCTGGCCCTCACCCGCCCACTAGTGTCCAGAGGCGTGTG 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGTGCCTGAACGGCCCCTGAGCCCTGGCCCTCACCCGCCCACTATGGTCCAGAGGCGTGTG 300

Qy   301 GGTGAGCGCCGCTGCTGCGGCACCGGAAAGCCTCTTGCTGTGGTCAACCTGCTAACCTT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGTGAGCGCCGCTGCTGCGGCACCGGAAAGCCCAGCTCTTGCTGTGGTCAACCTGCTAACCTT 360

Qy   361 TGGCCTGGAGGTGTTGTTTGGCCGCAGGCATCACCTATGTGCCGCCCTCTCGCTGCTGGAAGT 420

```
Db   361  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          TGGCCTGGAGGTGTGTTTGGCCGTGCCAGGCATCACCTACCGTGGGGCATTGGTCCAGTGCTGGGGCCTGGT  420

Qy   421  GGGGTGTAGAGAGGAAGAGAAGTTCATGACCATGGTCGTGGCATTGGTCCAGTGGTGGGCCTGGT  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  GGGGTGTAGAGAGGAAGAGAAGTTCATGACCATGGTCGTGGCATTGGTCCAGTGGTGGGCCTGGT  480

Qy   481  CTGTGCCCGCCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGGCTATGGCCGCCGCCCG  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  CTGTGCCCGCCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGGCTATGGCCGCCGCCCG  540

Qy   541  GCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTCATCCCAAGGGC  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GCCCTTCATCTGGGCACTGTCCTTGGGCATCCTGCTGAGCCTCTTCATCCCAAGGGC  600

Qy   601  CGGCTGGCTAGCAGGGCTGCTGTCTACCTCGGGGCTACCTGCTGGGGATCCCAGGCCCCTGGAGCTGGCACTGCTCAT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  CGGCTGGCTAGCAGGGCTGCTGTCTACCTCGGGGCTACCTGCTGGGGATCCCAGGCCCCTGGAGCTGGCACTGCTCAT  660

Qy   661  CCTGGGCGGTGGGGCTTCTGTGCCCAGGTGTCTTCACTCACTGGAGGCCCT  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  CCTGGGCGGTGGGGCTTCTGTGCCCAGGTGTCTTCACTCACTGGAGGCCCT  720

Qy   721  GCTCTCTGACCTCTTCCGGGGACCCACTGTCGCGGACCCGACCACTCTGTCTCTATGCCTTT  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GCTCTCTGACCTCTTCCGGGGACCCACTGTCGCGGACCCGACCACTCTGTCTCTATGCCTTT  780

Qy   781  CATGATCAGTCTTGGGGGCTGCCTGCCTACCTCCTGCCTGCCATTGACTGGACACCAG  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  CATGATCAGTCTTGGGGGCTGCCTGCCTACCTCCTGCCTGCCATTGACTGGACACCAG  840

Qy   841  TGCCCTGGCCCCCCTACCTGGGGCACCCAGGAGGAGTCCACCAGCTGTGCTCTCATCAT  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  TGCCCTGGCCCCCCTACCTGGGGCACCCAGGAGGAGTCCACCAGCTGTGCTCTCATCAT  900

Qy   901  CTTCCTCACCTGCTAGCAGCCCACACTGCTGGTGGCCAGGAGGCAGCCGCTGGGGCCCAC  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  CTTCCTCACCTGCTAGCAGCCCACACTGCTGGTGGCCAGGAGGCAGCCGCTGGGGCCCAC  960

Qy   961  CGAGCCAGCAGAGAGGGCTGTGTCGGGCCCCCCTTGTCGCGCCGCCCACTGTGCCGGGGC  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  CGAGCCAGCAGAGAGGGCTGTGTCGGGCCCCCCTTGTCGCGCCGCCCACTGTGCCGGGGC  1020

Qy  1021  CCGCTTCAGCGGATTCGGCGCCGGGAACCTGGGCGCCGCGCTTCCCCGGCCTGCGTGCTGCGG  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CCGCTTCAGCGGATTCGGCGCCGGGAACCTGGGCGCCGCGCTTCCCCGGCCTGCGTGCTGCGG  1080

Qy  1081  CATGCCCGCACCCTGCTTCGTGGCTGCTCTTCGTGCTGCAGCTGGATGGCCAGCTGGCACTCAT  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  CATGCCCGCACCCTGCTTCGTGGCTGCTCTTCGTGCTGCAGCTGGATGGCCAGCTGGCACTCAT  1140

Qy  1141  GACCTTCACGCGCTGTTTTACACGGCTGTGTGGGCGAGGCGTGTACCAGGCCTGCCCAG  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  GACCTTCACGCGCTGTTTTACACGGCTGTGTGGGCGAGGCGTGTACCAGGCCTGCCCAG  1200

Qy  1201  AGCTGAGCCGGGCACCGGGGCACCGAGGACCCGGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCT  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  AGCTGAGCCGGGCACCGGGGCACCGAGGACCCGGAGACACTATGATGAAGGCGTTCGGATGGGCAGCCT  1260

Qy  1261  GGGGCTGTTCCTGCAGTGCCCATCTCCCTGGTCTTCTCTCTGTCCTCGGACCGGCTGGT  1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  GGGGCTGTTCCTGCAGTGCCCATCTCCCTGGTCTTCTCTCTGTCCTCGGACCGGCTGGT  1320

Qy  1321  GCAGGATTCGGCACTCGAGCAGTCAGTGCCAGTTCCCTGTGGCTGCCAGCTTCCCTGTGGCTGC  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  GCAGGATTCGGCACTCGAGCAGTCAGTGCCAGTTCCCTGTGGCTGCCAGCTTCCCTGTGGCTGC  1380

Qy  1381  CGGTGCCACATGCCTGTCCCACAGTGTGGCCTGGTCGACAGTTCAGCCGCCCTCACCGCG  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  CGGTGCCACATGCCTGTCCCACAGTGTGGCCTGGTCGACAGTTCAGCCGCCCTCACCGGG  1440

Qy  1441  GTTCACCTTCTCAGCCCTGCTGAGATCTGCCCTCTACCACCGGGA
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  1441  GTTCACCTTCTCAGCCCTGCTGAGATCTGCCCTCTACCACCACTGGCCTCCCTCCTACCACCGGGA  1500

Qy  1501  GAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACAGCTGGAGGTGCTAGCAGTGAGGACACAG  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  GAAGCAGGTGTTCCTGCCCAAATACCGAGGGGACAGCTGGAGGTGCTAGCAGTGAGGACACAG  1560

Qy  1561  CCTGATGACCAGCTTCTGCCCAGGCCCTAAGCCTGGAGCTCCCTTCCTAATGGACACGT  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  CCTGATGACCAGCTTCTGCCCAGGCCCTAAGCCTGGAGCTCCCTTCCTAATGGACACGT  1620

Qy  1621  GGGTGCTGGAGGCAGTGGCCGTGCTCTGAGCCTCTTCATCCATGGCGGCTCTGCCTG  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  GGGTGCTGGAGGCAGTGGCCGTGCTCTGAGCCTCTTCATCCGCGGGGGCCTCTGCCTG  1680

Qy  1681  TGATGTCTCCGTACGTGTGGTGGTGGTGGTGAGCCTAGCCCTCCGTGAGTAGTATTTGACAAGAG  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  TGATGTCTCCGTACGTGTGGTGGTGGTGGTGAGCCGCCACCGAGGCCAGGTGGTTCCGGGCCG  1740

Qy  1741  GGGCATCTGCCTGGACCTCGCCATCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCC  1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  GGGCATCTGCCTGGACCTCGCCATCTGGATAGTGCCTTCCTGCTGTCCCAGGTGGCCCC  1800

Qy  1801  ATCCCTGTTTATGGGCTCCATTGTCCAGCTGCTCAGCCTCTGTCACTCGCTCTATATGTGTC  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  ATCCCTGTTTATGGGCTCCATTGTCCAGCTGCTCAGCCTCTGTCACTCGCTCTATATGTGTC  1860

Qy  1861  TGCCCGCAGGCCTGGCTCTGTGTCGCCATTACTTGCTCTACACAGGTAGTATTTGACAAGAG  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861  TGCCCGCAGGCCTGGCTCTGTGTCGCCATTACTTGCTCTACACAGGTAGTATTTGACAAGAG  1920

Qy  1921  CGACTTGCCAAATACTCAGGCTGTAGAAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCT  1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921  CGACTTGCCAAATACTCAGGCTGTAGAAAAACTTCCAGCACATTGGGGTGGAGGGCCTGCCT  1980

Qy  1981  CACTGGGTCCCAGCTCCTGTAGCCCATGGGGCTGCCGGCTGCCGCCGCCAGT  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  CACTGGGTCCCAGCTCCTGTAGCCCATGGGGCTGCCGGCTGCCGCCGCCAGT  2040

Qy  2041  TTCTGTTGCTGCCAAAGTAATGTGGCCTCTGCTGCTCCTCTCTGCTGCTGAGGTGCGTA  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  TTCTGTTGCTGCCAAAGTAATGTGGCCTCTGCTGCTCCTCTCTGCTGCTGAGGTGCGTA  2100

Qy  2101  GCTGCACAGCTGGGGCGTGGGGCGTGGGGCGTCCTCTCTCCCCAGTCTCTAGGGCTGCCTG  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  GCTGCACAGCTGGGGCGTGGGGCGTGGGGCGTCCTCTCTCTCCCAGTCTCTAGGGCTGCCTG  2160

Qy  2161  ACTGGAGGCCTTCCAAGGGGTTCAGTCTGGACTTATACAGGAGGCCAGAGAGGGCTCC  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  ACTGGAGGCCTTCCAAGGGGTTCAGTCTGGACTTATACAGGAGGCCAGAGAAGGGCTCC  2220

Qy  2221  ATGCACTGGAATGCGGGGGACTCTGCCAGTGGATTACCCAGGGTTAACAGCTAGC  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  ATGCACTGGAATGCGGGGGACTCTGCCAGTGGATTACCCAGGGTTAACAGCTAGC  2280

Qy  2281  CTCCTAGTTGAGACACACCTGAGGAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  CTCCTAGTTGAGACACACCTGAGGAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG  2340

Qy  2341  GTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTCTTGCATGGGAG  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  GTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTCTTGCATGGGAG  2400

Qy  2401  TTTCTAGGATGAAAACACTCCTCCATGGGATTTGAACATATGGACTTATTTGTAGGGGAAGA  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TTTCTAGGATGAAAACACTCCTCCATGGGATTTGAACATATGGACTTATTTGTAGGGGAAGA  2460

Qy  2461  GTCCTGAGGGGCAACACAAGAACCAGGTCCCCTCAGCCCACCAGCACTGTCTTTTTTGCT  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GTCCTGAGGGGCAACACAAGAACCAGGTCCCCTCAGCCCACCAGCACTGTCTTTTTTGCT  2520

Qy  2521  GATCCACCCCCTCTTACCTTTTATCAGGATGTGGCCTGTTGTCCTTCTGTTGTCCATCA  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  GATCCACCCCCTCTTACCTTTTATCAGGATGTGGCCTGTTGTCCTTCTGTTGTCCATCA  2580
```

```
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
DR   P-PSDB; W71869.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 79-80; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence  3410 BP;        667 A;     1014 C;      945 G;      783 T;

     Query Match             100.0%;      Score 3409.6;  DB 1;   Length 3410;
     Best Local Similarity   100.0%;      Pred. No. 0;
     Matches 3410;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy         1 GGGAACCAGCCTGCACGCGCTGGCTCGGGTGACACGCGCCGGCCGCCTCGGCCAGGATCTGA 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 GGGAACCAGCCTGCACGCGCTGGCTCGGGTGACACGCGCCGGCCGCCTCGGCCAGGATCTGA 60

Qy        61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACCAGCAGGAGTGTTGAGCATGGGCTGAG 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GTGATGAGACGTGTCCCACTGAGGTGCCCCACCAGCAGGAGTGTTGAGCATGGGCTGAG 120

Qy       121 AAGCTGGACCGGCACCAAAGGGCTGGCAGAAATGGGCGGCCTGGCTGATTCCTAGGCAGTT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AAGCTGGACCGGCACCAAAGGGCTGGCAGAAATGGGCGGCCTGGCTGATTCCTAGGCAGTT 180

Qy       181 GGCGGCAAGGAGGAGGCCGAGCTTCTGGACAGACCGAGCAGAAGCAGTTCTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GGCGGCAAGGAGGAGGCCGAGCTTCTGGACAGACCGAGCAGAAGCAGTTCTG 240

Qy       241 GAGTGCCTGAACGGCCCCCTGAGCCCTACCCGCCTGGCCCACTATGGTCCAGAGGCGTG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GAGTGCCTGAACGGCCCCCTGAGCCCTACCCGCCTGGCCCACTATGGTCCAGAGGCGTG 300

Qy       301 GGTGAGCCGCCGCTGCGCGCACCCGGAAAGCCCAGCTCTTGCTGCTGGTCAACCTGCTAACCTT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 GGTGAGCCGCCGCTGCGCGCACCCGGAAAGCCCAGCTCTTGCTGCTGGTCAACCTGCTAACCTT 360

Qy       361 TGGCCTGGAGTGTGTTGGCGCAGGCATCACCTATGTGCGCCTCGTGCTGCTGAAGT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TGGCCTGGAGTGTGTTGGCGCAGGCATCACCTATGTGCGCCTCGTGCTGCTGGAAGT 420

Qy       421 GGGGGTAGGAGGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGGGCTGGGTCCTGGT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GGGGGTAGGAGGAAGTTCATGACCATGGTGCTGGGCATTGGTCCAGTGGGCTGGGTCCTGGT 480

Qy       481 CTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCG 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CTGTGTCCCGCTCCTAGGCTCAGCCAGTGACCACTGGCGTGGACGCTATGGCCGCCGCCG 540

Qy       541 GCCCTTCATCTGGGCACTGCTTCCTGTGCCCGATCCTGCTGAGCTGGCACTGCTCAT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 GCCCTTCATCTGGGCACTGCTTCCTGTGCCCGATCCTGCTGAGCTGGCACTGCTCAT 600

Qy       601 CGGCTGGCTAGCAGGCTGGGGCGTGGGGCTGTCGGGACTTCTGTGGCCAGTGCTTCACTCCACTGGAGCCCT 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CGGCTGGCTAGCAGGCTGGGGCGTGGGGCTGTCGGGACTTCTGTGGCCAGTGCTTCACTCCACTGGAGCCCT 660

Qy       661 CCTGGGCGTGGGGCTGTCGGGACTTCTGTGGCCAGTGCTTCACTCCACTGGAGCCCT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CCTGGGCGTGGGGCTGTCGGGACTTCTGTGGCCAGTGCTTCACTCCACTGGAGCCCT 720

Qy       721 GCTCTCTGACCCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCCTT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 GCTCTCTGACCCTCTTCCGGGACCCGGACCACTGTCGCCAGGCCTACTCTGTCTATGCCTT 780

Qy       781 CATGATCAGTTCTTGGGGCTGCCTGGGGCTGCTCCTGCCATTGACTGGGACACCAG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy      2581 CAGAGACACAGGCATTTAAATATTATTTATTTTAACAAAGTAGAAGGGAATCCAT 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 CAGAGACACAGGCATTTAAATATTATTTATTTTAACAAAGTAGAAGGGAATCCAT 2640

Qy      2641 TGCTAGCTTTTCTGTGTTGGTGTCTAAATATTGGGTAGGGTGGGGATCCCAACAATCA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 TGCTAGCTTTTCTGTGTTGGTGTCTAAATATTGGGTAGGGTGGGGATCCCAACAATCA 2700

Qy      2701 GGTCCCCTGAGATAGCTGGTCATGGCGTGATCATTGCCAGAATCTTCTTCTCCTGGGGT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 GGTCCCCTGAGATAGCTGGTCATGGCGTGATCATTGCCAGAATCTTCTTCTCCTGGGGT 2760

Qy      2761 CTGGCCCCCAAAAATGCCTAACCCAGGACCTTGGAAATTCTACTCACTCCCAAATGATAAT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 CTGGCCCCCAAAAATGCCTAACCCAGGACCTTGGAAATTCTACTCACTCCCAAATGATAAT 2820

Qy      2821 TCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 TCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAAGGAAGGTAGAGGGTGGGGCTTCAGGT 2880

Qy      2881 CTCAACGGCTTCCCTAACCACCCCCTCTTCTCTTGCCCCAGCCTGGTTCCCCACTTCCA 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 CTCAACGGCTTCCCTAACCACCCCCTCTTCTCTTGCCCCAGCCTGGTTCCCCACTTCCA 2940

Qy      2941 CTCCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTCCCCTACC 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 CTCCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTCCCCTACC 3000

Qy      3001 CCCAACTTTCCCCTACCCCCAACTTCCCACACCTCCACAACCCTGTTTGGAGCTACT 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 CCCAACTTTCCCCTACCCCCAACTTCCCACACCTCCACAACCCTGTTTGGAGCTACT 3060

Qy      3061 GCAGGACCAGAAGCACAAAGTGCCGGTTCCCAAGCCCTTGTGTCCATCTCCAGCCCCCAGAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 GCAGGACCAGAAGCACAAAGTGCCGGTTCCCAAGCCCTTGTGTCCATCTCCAGCCCCCAGAGT 3120

Qy      3121 ATATCTGTCTTGGGGAATCTCACAGAAACTCAGGAGGCACCCCCTGCCCTGAGCTAAGG 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 ATATCTGTCTTGGGGAATCTCACAGAAACTCAGGAGGCACCCCCTGCCCTGAGCTAAGG 3180

Qy      3181 GAGGTCTTATCTCTCAGGGGGGGGTTTAAGTGCCGTTTGCCAATATGTGTCTTATTTATT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 GAGGTCTTATCTCTCAGGGGGGGGTTTAAGTGCCGTTTGCCAATATGTGTCTTATTTATT 3240

Qy      3241 TAGGGGGGTGAATATTTTATACTGTGAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 TAGGGGGGTGAATATTTTATACTGTGAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA 3300

Qy      3301 AAATTAAAGGCTTTCTTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 AAATTAAAGGCTTTCTTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3360

Db      3361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3410

Db      3361 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAA 3410
```

```
RESULT   2
V61201
ID   V61201 standard; cDNA; 3410 BP.
AC   V61201;
DT   06-JAN-1999  (first entry)
DE   Full length cDNA sequence of prostate tumour clone Li-12.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   25-FEB-1998; US-020956.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
```

```
Db    781  CATGATCAGTCTGGGGGCTGCCTGGCCTACCTCCTGGCTGCCATTGACTGGGACACCAG  840

Qy    841  TGCCCTGGCCCCTACCTGGCACCCAGGAGGAGTCCTCTTTGGCCTGCTCACCCTCAT  900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  TGCCCTGGCCCCTACCTGGCACCCAGGAGGAGTCCCTCTTTGGCCTGCTCACCCTCAT  900

Qy    901  CTTCCTCACCTGCGTAGCAGCCACACTGCTGGTGCTGGCAGGAGGCGCTGGGCCCCAC  960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  CTTCCTCACCTGCGTAGCAGCCACACTGCTGGTGCTGGCAGGAGGCGCTGGGCCCCAC  960

Qy    961  CGAGCCAGCAGAGGGCTGTCGGCCCCACTCGTGTCGCCCCACTGCGTGTCCATGCGGGC  1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  CGAGCCAGCAGAGGGCTGTCGGCCCCACTCGTGTCGCCCCACTGCGTGTCCATGCGGGC  1020

Qy   1021  CCGCTTGGCTTTCCGGAACCTGGGCGCGCTGTCCCCGGCTGCACCAGCTGTGCTGCCG  1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  CCGCTTGGCTTTCCGGAACCTGGGCGCGCTGTCCCCGGCTGCACCAGCTGTGCTGCCG  1080

Qy   1081  CATGCCCCGCACCCTGCGCCGGCCCGCTCTTCGTGGCTGTGCAGCTGGATGGCACTCAT  1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  CATGCCCCGCACCCTGCGCCGGCCCGCTCTTCGTGGCTGTGCAGCTGGATGGCACTCAT  1140

Qy   1141  GACCTTCACGCTGTGTTTACACGGATTTCGTGGCGGAGGGGCTGTACCAGGGCTGCCCAG  1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  GACCTTCACGCTGTGTTTACACGGATTTCGTGGCGGAGGGGCTGTACCAGGGCTGCCCAG  1200

Qy   1201  AGCTGAGCGGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGAATGGGCAGCCT  1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  AGCTGAGCGGGGCACCGAGGCCCGGAGACACTATGATGAAGGCGTTCGGAATGGGCAGCCT  1260

Qy   1261  GGGCTGTTCCTGCAGTGCCGCCATCTCCCTGTCTCTCTCGGTCATGATGGGACCGGCTGGT  1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  GGGCTGTTCCTGCAGTGCCGCCATCTCCCTGTCTCTCTCGGTCATGATGGGACCGGCTGGT  1320

Qy   1321  GCAGCGGATTCGGCACTGCGAGCAGTCTATTTGCCGGTGTGGCCAGCTTTCCCTGTGGCTGC  1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  GCAGCGGATTCGGCACTGCGAGCAGTCTATTTGCCGGTGTGGCCAGCTTTCCCTGTGGCTGC  1380

Qy   1381  CGGTGCCACATGCTTGCTGGCCGTGTGCCGGTGACAGCTTCAGCGCGCCCTCACCGG  1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  CGGTGCCACATGCTTGCTGGCCGTGTGCCGGTGACAGCTTCAGCGCGCCCTCACCGG  1440

Qy   1441  GTTCACCTTCTCAGCCCTGCCCTGCAGATCCTGCCCTCACACACTGGCCTCTCACCACCGGGA  1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  GTTCACCTTCTCAGCCCTGCCCTGCAGATCCTGCCCTCACACACTGGCCTCTCACCACCGGGA  1500

Qy   1501  GAAGCAGGTGTTCTCCGTACGTGTGTGGGTGAGCCCCACCGGGTGGTTAGCAGTGAGGACAG  1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  GAAGCAGGTGTTCTCCGTACGTGTGTGGGTGAGCCCCACCGGGTGGTTAGCAGTGAGGACAG  1560

Qy   1561  CCTGATGACCAGCTTCCTGCCAAAGCCCTAAGCCTGGAGCTCCCTAAGCTCCCTAAATGGACACGT  1620
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  CCTGATGACCAGCTTCCTGCCAAAGCCCTAAGCCTGGAGCTCCCTAAGCTCCCTAAATGGACACGT  1620

Qy   1621  GGGTGCTGAGGCAGTGGGCCTGCTCCCACCTCCACCGGCCTCGCCGGGGCCTCTGCCTG  1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  GGGTGCTGAGGCAGTGGGCCTGCTCCCACCTCCACCGGCCTCGCCGGGGCCTCTGCCTG  1680

Qy   1681  TGATGTCTCCGTACGTGTGTGGTGGGTGAGCCCCACCGGGTGGTTCCCGGGCCG  1740
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  TGATGTCTCCGTACGTGTGTGGTGGGTGAGCCCCACCGGGTGGTTCCCGGGCCG  1740

Qy   1741  GGGCATCTGCCTGACCCTGGCCATCTGCCGACCTGGATAGTGCCTAAGCTGTCTGCCAGGTGGCCCCC  1800
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  GGGCATCTGCCTGACCCTGGCCATCTGCCGACCTGGATAGTGCCTAAGCTGTCTGCCAGGTGGCCCCC  1800

Qy   1801  ATCCCTGTTTATGGCTCACTGCTCCATTGCCTCACGTCGTCACTGCCTATATGGTGTC  1860
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  ATCCCTGTTTATGGCTCACTGCTCCATTGCCTCACGTCGTCACTGCCTATATGGTGTC  1860

Qy   1861  TGCCGCAGGCCTGGCTCGTCGCTGCTCGCCCATTTACTTTGCTCACACAGTAGTATTTGACAAGAG  1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  TGCCGCAGGCCTGGCTCGTCGCTGCTCGCCCATTTACTTTGCTCACACAGTAGTATTTGACAAGAG  1920
```

```
Qy   1921  CGACTTGGCCAAATACTCAGCGTGAGAAAACTTCAGCACATTGGGTGGAGGGCCTGCCT  1980
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  CGACTTGGCCAAATACTCAGCGTGAGAAAACTTCAGCACATTGGGTGGAGGGCCTGCCT  1980

Qy   1981  CACTGGGTCCCAGCTCCCGCTCCTGTAGCCCCATGGGCGGCTGCCGGGCTGCCGGGCTGCCGCCAGT  2040
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  CACTGGGTCCCAGCTCCCGCTCCTGTAGCCCCATGGGCGGCTGCCGGGCTGCCGGGCTGCCGCCAGT  2040

Qy   2041  TTCTGTTGCTGCCAAAGTAATGTGCCTCTGCTGCTGCCACCCTGTTGCTGCTGAGTGCGTA  2100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  TTCTGTTGCTGCCAAAGTAATGTGCCTCTGCTGCTGCCACCCTGTTGCTGCTGAGTGCGTA  2100

Qy   2101  GCTGCACAGTGGGGGCTGGGGCGCTGGGGCGTCCCTCTCCCCAGTCTCTAGGGCTCTGCCTG  2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  GCTGCACAGTGGGGGCTGGGGCGCTGGGGCGTCCCTCTCCCCAGTCTCTAGGGCTCTGCCTG  2160

Qy   2161  ACTGGAGGCCTTCCAAGGGGTTTCAGTCTGGACTTATACAGGGAGGCCAGAAGGGCTCC  2220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161  ACTGGAGGCCTTCCAAGGGGTTTCAGTCTGGACTTATACAGGGAGGCCAGAAGGGCTCC  2220

Qy   2221  ATGCACTGGAATGCGGGACTCTGCAGGTGATTACCCAGGGTTAACAGCTAGC  2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  ATGCACTGGAATGCGGGACTCTGCAGGTGATTACCCAGGGTTAACAGCTAGC  2280

Qy   2281  CTCCTAGTTGAGACACACCTAGAGAGAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG  2340
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  CTCCTAGTTGAGACACACCTAGAGAGAGGGTTTTGGGAGCTGAATAAACTCAGTCACCTG  2340

Qy   2341  GTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTCTTGCATGGGAG  2400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  GTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTCTTGCATGGGAG  2400

Qy   2401  TTTCTAGGATGAAAACACTCCTCCATGGGATTTGAACACATATGACTTATTTGTTAGGGGAAGA  2460
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401  TTTCTAGGATGAAAACACTCCTCCATGGGATTTGAACACATATGACTTATTTGTTAGGGGAAGA  2460

Qy   2461  GTCCTGAGGGGCAACACAAGAACCAGGTCCCCTGGCCTGTGGTCCTTCGTGTTGCCATCA  2520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461  GTCCTGAGGGGCAACACAAGAACCAGGTCCCCTGGCCTGTGGTCCTTCGTGTTGCCATCA  2520

Qy   2521  GATCCACCCCCCTCTTACCTTTATCAGGATGTGGCCTGTGTTGGTCCTGTGTTGCCATCA  2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521  GATCCACCCCCCTCTTACCTTTTATCAGGATGTGGCCTGTGTTGGTCCTGTGTTGCCATCA  2580

Qy   2581  CAGAGACACAGGCATTTAAATATTAAAATATTTAACTAGTTAATTTATTTAACAAAGTAGAAGGGAATCCAT  2640
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581  CAGAGACACAGGCATTTAAATATTAAAATATTTAACTAGTTAATTTATTTAACAAAGTAGAAGGGAATCCAT  2640

Qy   2641  TGCTAGCTTTCTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCAACAATCA  2700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641  TGCTAGCTTTCTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCAACAATCA  2700

Qy   2701  GGTCCCCTGAGATAGCTGGTCATTGGGCTGATCATTGCCAGAATCTTCTTCCTCTGGGGT  2760
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701  GGTCCCCTGAGATAGCTGGTCATTGGGCTGATCATTGCCAGAATCTTCTTCCTCTGGGGT  2760

Qy   2761  CTGGCCCCCAAAATGCCTAACCCAGGACACCTTGGAAATTCTACTACTCATCCCAAATGATAAT  2820
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761  CTGGCCCCCAAAATGCCTAACCCAGGACACCTTGGAAATTCTACTACTCATCCCAAATGATAAT  2820

Qy   2821  TCCAAATGCTTACCCAAGGTAGGGTTGTTGAAGGAAGTAGAGGGCTTCAGGT  2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  TCCAAATGCTTACCCAAGGTAGGGTTGTTGAAGGAAGTAGAGGGCTTCAGGT  2880

Qy   2881  CTCAACGGCTTCCTAACCACCCCTCTTCCTGGCCCCCAGCCTGTTCCCCCACTTCCA  2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881  CTCAACGGCTTCCTAACCACCCCTCTTCCTGGCCCCCAGCCTGTTCCCCCACTTCCA  2940

Qy   2941  CTCCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTTCCCCTACC  3000
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  CTCCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTTCCCCTACC  3000
```

```
Qy  3001  CCCAACTTCCCTACCCCAACTTTCCCACCCAGCTCCACAACCCTGTTGGGAGCTACT  3060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  CCCAACTTCCCTACCCCAACTTTCCCACCCAGCTCCACAACCCTGTTGGGAGCTACT  3060

Qy  3061  GCAGGACCAGAAGCACAAAGTGCGGTTTCCAAGCCTTTGTCCATCTCAGCCCCAGAGT  3120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3061  GCAGGACCAGAAGCACAAAGTGCGGTTTCCAAGCCTTTGTCCATCTCAGCCCCAGAGT  3120

Qy  3121  ATATCTGTGTTGCTTGGGGAATCTCACCACAGAAACTCAGGAGGCACCCCCTGCCTAAGG  3180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3121  ATATCTGTGTTGCTTGGGGAATCTCACCACAGAAACTCAGGAGGCACCCCCTGCCTAAGG  3180

Qy  3181  GAGGTCTTATCTCTCAGGGGGGTTAAGTGCCGTTTGCAATAGTGTCGTCTTATTTATT  3240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3181  GAGGTCTTATCTCTCAGGGGGGTTAAGTGCCGTTTGCAATAGTGTCGTCTTATTTATT  3240

Qy  3241  TAGCGGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA  3300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3241  TAGCGGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACA  3300

Qy  3301  AAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  AAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  3360

Qy  3361  AAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAA  3410
          |||||||||||||||||||||||||||||||||||||||
Db  3361  AAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAA  3410
```

```
Query Match                     62.7%;  Score 2136.4;  DB 1;  Length 2152;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 2149;  Conservative   0;  Mismatches   0;  Indels   2;  Gaps   1;

Qy  1177  GGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGCACCGAGGGCCGGAGACACTATGA  1236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGCACCGAGGGCCGGAGACACTATGA  60

Qy  1237  TGAAGGCGTTCGGATGGGCAGCCTGGGCGTGTTCCTGCAGTGCGCGCCATCTCCTGTCTT  1296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TGAAGGCGTTCGGATGGGCAGCCTGGGCGTGTTCCTGCAGTGCGCGCCATCTCCTGTCTT  120

Qy  1297  CTCTCTGGTCATGACCGGCTGGTGCCGGTGCCACTCGAGCAGTCATTGGCCAG  1356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CTCTCTGGTCATGACCGGCTGGTGCCGGTGCCACTCGAGCAGTCATTGGCCAG  180

Qy  1357  TGTGGAGCTTCCCTGTGCGTGCCGGTGCCACATCCTGTCCCACAGTGTGGCCGTGGT  1416
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  TGTGGAGCTTCCCTGTGCGTGCCGGTGCCACATCCTGTCCCACAGTGTGGCCGTGGT  240

Qy  1417  GACAGCTTCAGCCGCCCTCACCGGTTCACCTTCTCAGCCTGCGAGATCCTGCCTACAC  1476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GACAGCTTCAGCCGCCCTCACCGGTTCACCTTCTCAGCCTGCGAGATCCTGCCTACAC  300

Qy  1477  ACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCAAATACCGAGGGGACAC  1536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  ACTGGCCTCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCAAATACCGAGGGGACAC  360

Qy  1537  TGGAGTGCTGCTAGCAGTGAGGACAGCAGCCTGATGACCAGCTTCCTGCTGCCAGGCCTGG  1596
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  TGGAGTGCTGCTAGCAGTGAGGACAGCAGCCTGATGACCAGCTTCCTGCTGCCAGGCCTGG  420

Qy  1597  AGCTCCTTCCTTCCTAATGACACGTGGCTGTGAGGCAGTGGCCTGCTCCCCACCTCCACC  1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  AGCTCCTTCCTTCCTAATGACACGTGGCTGTGAGGCAGTGGCCTGCTCCCCACCTCCACC  480

Qy  1657  CGGCGTCTGCGGGCCCTCTGCGCTCCGTGATGTCCGTGGTGGTGGTGAGCCCAC  1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  CGGCGTCTGCGGGCCCTCTGCGCTCCGTGATGTCCGTGGTGGTGGTGAGCCCAC  540

Qy  1717  CGAGGCCAGGTGGTTCCGGCCAGGTGGCATCGCCTGCCGACCTCGGATATGTGC  1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  CGAGGCCAGGTGGTTCCGGCCAGGTGGCATCGCCTGCCGACCTCGGATAGTGC  600

Qy  1777  CTTCCTGCTGTCGCCCAGGTGGCCTCCAGTTATGGGCTCCAGTCTCAGCTCAGCCA  1836
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  CTTCCTGCTGTCGCCCAGGTGGCCTCCAGTTATGGGCTCCAGTCTCAGCTCAGCCA  660

Qy  1837  GTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCGTCGGTCGCCATTACTTTGC  1896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GTCTGTCACTGCCTATATGGTGTCTGCCGCAGGCGTCGGTCGCCATTACTTTGC  720

Qy  1897  TACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGGGTAGAAAACTTCCAG  1956
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  TACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGGGTAGAAAACTTCCAG  780

Qy  1957  CACATTGGGGTGGAGGGCTCCACTGGGCTCCCAGTTCCCCAT  2016
          ||||||||||||||||||||||||||||||||||||||||||
Db   781  CACATTGGGGTGGAGGGCCTGGCCCTGGGCTCCCAGTTCCCCAT  840

Qy  2017  GGGGGTGCCGGGCTGCCGCGCAGTTCTGTGTCGCCAAAGTAGTAATGTGGCTCTCTGCTGC  2076
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  GGGGGTGCCGGGCTGCCGCGCAGTTCTGTGTCGCCAAAGTAATGTGGCTCTCTGCTGC  900

Qy  2077  CACCCTGTGCTGCTGAGGTGCGTAGCTGCACAGCTGGGGGCTGGGGCGTCCTCCTCTCT  2136
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  CACCCTGTGCTGCTGAGGTGCGTAGCTGCACAGCTGGGGGCTGGGGCGTCCTCCTCTCT  960

Qy  2137  CTCCCCAGTCTCTAGGGCTCTGACTGAGGCCTTCCAAGGGGGTTCAGTCTGGACTT  2196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  CTCCCCAGTCTCTAGGGCTCTGACTGAGGCCTTCCAAGGGGGTTCAGTCTGGACTT  1020

Qy  2197  ATACAGGGAGGCCAGAAGGGCTCCATGCACTGGAATGCGGGACTTCTGCAGGTGGATTAC  2256
```

```
RESULT   3
V71181
ID   V71181 standard; cDNA; 2152 BP.
AC   V71181;
DT   12-FEB-1999  (first entry)
DE   Consensus sequence of the PS108 gene derived from overlapping clones.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Synthetic.
OS   Homo sapiens.
FH   Key            Location/Qualifiers
FT   CDS            2..769
FT                  /tag=  a
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI: 99-034731/03.
DR   P-PSDB; W85068.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1: Fig 1A-E; 122pp; English.
CC   The present sequence represents the consensus sequence for a PS108
CC   gene. The sequence is derived from overlapping clones v71166-79. The
CC   clone sequences are PS108 gene-specific. They are used in the method
CC   of the invention. The specification describes a method for detecting the
CC   presence of a target PS108 polynucleotide in a test sample. The method
CC   comprises contacting the test sample with at least 1 PS108-specific
CC   polynucleotide or complement, and detecting the presence of the target
CC   PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence    2152 BP;   419 A;   622 C;   569 G;   542 T;
```

Db   2101 ATCAGAGTATAATGTTTATGTGACAAAATTAAAGGCTTTCTTATATGTTTA 2152

RESULT   4
V71180
ID   V71180 standard; cDNA; 2143 BP.
AC   V71180;
DT   12-FEB-1999 (first entry)
DE   Clone 1711346IH, the PS108 gene contig full length sequence.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-P; 122pp; English.
CC   The present sequence represents the full length contig of the PS108
CC   gene, as represented by clone 1711346IH. This clone is the contig of
CC   overlapping clones V71166-79. The clone sequences are PS108
CC   gene-specific. They are used in the method of the invention. The
CC   specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence   2143 BP;   418 A;   563 G;   621 C;   541 T;

Query Match   62.0%; Score 2114.8; DB 1; Length 2143;
Best Local Similarity   99.8%; Pred. No. 0;
Matches 2139;   Conservative   0;   Mismatches   0;   Indels   2;   Gaps   3;

Qy   1185 ACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACTATGATGAGGCG 1244
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACCAGGGCGTGCCCAGAGCTGAGCCGGGCACCGAGGCCCGGAGACTATGATGAGGCG   60

Qy   1245 TTCGGATGGGCAGCCTGGGGCTGTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGG 1304
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTCGGATGGGCAGCCTGGGGCTGTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTCTGG   120

Qy   1305 TCATGGACCGGCTGGTGCAGCGATTCGGCCACTCGAGCAGTCTATTGGCCAGTGTGGCAAG 1364
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCATGGACCGGCTGGTGCAGCGATTCGGCCACTCGAGCAGTCTATTGGCCAGTGTGGCAAG   180

Qy   1365 CTTTCCCTGTGGCTGCCGGTGCCGGTGCCACATGCTTGCCCACAGTGTGGCCGTGGTGACAGCTT 1424
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTTTCCCTGTGGCTGCCGGTGCCGGTGCCACATGCTTGCCCACAGTGTGGCCGTGGTGACAGCTC   240

Qy   1425 CAGCCGCCTCACCGGCTTCACCTTCGCAGATCGATCCCCTGCGAGATCGATCCCTACACACTGGCCT 1484
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAGCCGCCTCACCGGCTTCACCTTCGCAGATCGATCCCCTGCGAGATCGATCCCTACACACTGGCCCT   300

Qy   1485 CCCTCTACCACCGGGAGAAGCACAGTGTTCCTGCCCAAATACCGAGGGGACACTGGAGTG 1544
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCCTCTACCACCGGGAGAAGCAGCAGTGTTCCTGCCCAAATACCGAGGGGACACTGGAGTG   360

Qy   1545 CTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGCCAGCTTCCTGCGAGCTCCCT 1604
           ||||||||||||||||||||||||||||||||||||||||||

```
Db   361  CTAGCAGTGAGGACAGCCTGATGACGCTTCCTGCGCAGGCCCTAAGCCTTGAGCTCCCT  420

Qy  1605  TCCCTAATGACACGTGGGTGCTGAGGCAGTGGCCTGGCTCCTCCACCTCCACCCGCTCT  1664
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  TCCCTAATGGACACGTGGGTGCTGAGGCAGTGGCCTGGCTCCTCCACCTCCACCCGCTGC  480

Qy  1665  GCGGGGCCCTCGCCTGTGATGTCTCCGTACGTGTGGTGGTGAGCACCGAGGCCA  1724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GCGGGGCCCTCGCCTGTGATGTCTCCGTACGTGTGGTGGTGAGCACCGAGGCCA  540

Qy  1725  GGGTGGTTCCGGCGCCATCGCTGGACCTGCCATCCTGGATAGTGCCTTCCTGC  1784
          |||||||||||||||   ||||||||||||||||||||||||||||||||||
Db   541  GGGTGGTTCCGGCGCCGGGCCATCGCTGGACCTGCCATCCTGGATAGTGCCTTCCTGC  600

Qy  1785  TGTCCCAGGTGGCCCATCCCTGTTTATGGGCTCCATTGTCAGCTCAGCCAGTCTGTCA  1844
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   601  TGTCCCAGGTGGCCCCATCCCTGTTTATGGGCTCCATTGTCAGCTCAGCCAGTCTGTCA  660

Qy  1845  CTGCCTATATGGTGTCTGCGCAGGCCTGGGTCGTCGTGCCATTTACTTTGCTACACAGG  1904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  CTGCCTATATGGTGTCTGCGCAGGCCTGGGTCGTCGTGCCATTTACTTTGCTACACAGG  720

Qy  1905  TAGTATTTGACAAGAGCGACTTGGCCAAATACTCAGCGTAGAAAACTTCAGCACATTGG  1964
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   721  TAGTATTTGACAAGAGCCGACTTGGCCAAATACTCAGCGTAGAAAACTTCAGCACATTGG  780

Qy  1965  GGTGGAGGGCCTGCCTCACTGGGTCCCAGCTCCTGTTAGCCCCGCTCGTGTGCCGGGCTGC  2024
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  GGTGGAGGGCCTGCCTCACTGGGTCCCAGCTCCTGTTAGCCCCGCTCGTGTGCCGGGCTGC  840

Qy  2025  CGGGCTGCTGGCCGCCCAGTTTCTGTGTGCCAAAGTAATGTGGCTCTCTGCTCCACCCTGT  2084
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  CGGGCTGCTGGCCGCCCAGTTTCTGTGTGCCAAAGTAATGTGGCTCTCTGCTCCACCCTGT  900

Qy  2085  GCTGCTGAGGTGCGTAGCTGCACAGCTGGGCGTGGGGCGTCCCTCTCCTCCCCAG  2144
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  GCTGCTGAGGTGCGTAGCTGCACAGCTGGGCGTGGGGCGTCCCTCTCCTCCCCAG  960

Qy  2145  TCTCTAGGGGCTTCAGTCTGGAGGGCCTTCCAAGGGGGTTCAGTCTGGACTTATACCAGGG  2204
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  TCTCTAGGGGCTTCAGTCTGGAGGGCCTTCCAAGGGGGTTCAGTCTGGACTTATACCAGGG  1020

Qy  2205  AGGGCCAGAAGGGCTCCATGCAGTTCAGAATGCCGGGACTCTGCCAGGTGGATTACCCAGGAA  2264
          ||||||||||||||||||||||||||||||||| |||||||||||  ||||||||||||||||
Db  1021  AGGGCCAGAAGGGCTCCATGCAGTTCAGAATGCC-GGGACTCTGCGAGTGGATTACCCAGGCTC  1079

Qy  2265  AGGGTAACAGCTAGCCTCCAGTTGAGACAGCACCTAGAGAGGAGGGTTTTGGGAGCTGAA  2324
          |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1080  AGGGTTAACAGCTAGCCTCCAGTTGAGACAGCACCTAGAGAGGAGGGTTTTGGGAGCTGAA  1139

Qy  2325  TAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGT  2384
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1140  TAAACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGT  1199

Qy  2385  AGCTCTTGCATGGGAGTTTCTAGGATGAAACACTCCTTCCATGGGATTTGAACATATG--A  2442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db  1200  AGCTCTTGCATGGGAGTTTCTAGGATGAAACACTCCTTCCATGGGATTTGAACATATGAAA  1259

Qy  2443  CTTATTTGTAGGGGAAGAGTCCTGAGGGGCAACACAGAACCCAGTCCCCTCAGCCCA  2502
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1260  GTTATTTGTAGGGGAAGAGTCCTGAGGGGCAACACAGAACCCAGTCCCCTCAGCCCA  1319

Qy  2503  CAGCACTGTCTTTTTGCTGATCCACCCCCCTTTATCAGGATGTGGCCTGTTG  2562
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1320  CAGCACTGTCTTTTTGCTGATCCACCCCCCTTTATCAGGATGTGGCCTGTTG  1379

Qy  2563  GTCCTTCTGTTGCCATCACAGAGACACAGGCATTTAAATATTAAATATTATTTATTAAACA  2622
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1380  GTCCTTCTGTTGCCATCACAGAGACACAGGCATTTAAATATTAAATATTATTTATTAAACA  1439

Qy  2623  AAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGGTCTAATAATTTGGGTAGGGTG  2682
          ||||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||||
Db  1440  AAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGGTGTCTAATAATTTGGGTAGGGTG  1499


Qy  2683  GGGGATCCCACAATCAGGTCCCTGAGATAGTCTGGTCATTGGGCTGATCATTGCCAGA  2742
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1500  GGGGATCCCACAATCAGGTCCCTGAGATAGTCTGGTCATTGGGCTGATCATTGCCAGA  1559

Qy  2743  ATCTTCTCTCCTCGGGGTCTGGCCCCCCAAAATGCTAACCCAGGACCTTGAAATTCTA  2802
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1560  ATCTTCTCTCCTCGGGGTCTGGCCCCCCAAAATGCTAACCCAGGACCTTGAAATTCTA  1619

Qy  2803  CTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAAGGAAGTA  2862
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1620  CTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAAGGAAGTA  1679

Qy  2863  GAGGGTGGGCTTCAGTCTCAACGGCTTCCCTAACCACCCCTCTTCTTGCCCCAGCC  2922
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1680  GAGGGTGGGCTTCAGTCTCAACGGCTTCCCTAACCACCCCTCTTCTTGCCCCAGCC  1739

Qy  2923  TGGTTCCCCCACTTCCACTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTG  2982
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1740  TGGTTCCCCCACTTCCACTCCCTCTACTCTCTCTAGGACTGGGCTGATGAAGGCACTG  1799

Qy  2983  CCCAAAATTCCCCTACCCCCAACTTCCCCTACCCCCAACTTCCCCACCAGTCCACA  3042
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1800  CCCAAAATTCCCCTACCCCCAACTTCCCCTACCCCCAACTTCCCCACCAGTCCACA  1859

Qy  3043  ACCCTGTTGGAGCTACTGCGAGGACGCAGGAAGCACAAAGTGCGGTTCGCGGTTTGTC  3102
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1860  ACCCTGTTGGAGCTACTGCGAGGACGCAGGAAGCACAAAGTGCGGTTCGCGGTTTGTC  1919

Qy  3103  CATCTCAGCCCCCAGAGTATATCTGCTGGGGAATCTCACACAGAAACTCAGGAGCAC  3162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1920  CATCTCAGCCCCCAGAGTATATCTGCTGGGGAATCTCACACAGAAACTCAGGAGCAC  1979

Qy  3163  CCCCTGCCTGAGCTAAGGAGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT  3222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1980  CCCCTGCCTGAGCTAAGGAGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT  2039

Qy  3223  AATGTCGTCTTATTATTTAGGCGGGTGAAATATTTTATACTGTAAGTGAGCAATCAGAGT  3282
          |||||||||||||||||| ||||||||| |||||||||||||||||||||||||||||||
Db  2040  AATGTCGTCTTATTATTTTTAGGCGGGTGAATATTTTATACTGTAAGTGAGCAATCAGAGT  2099

Qy  3283  ATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA  3326
          ||||||||||||||||||||||||||||||||||||||||||||
Db  2100  ATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA  2143
```

RESULT   5
V58487
ID   V58487 standard; cDNA; 789 BP.
AC   V58487;
DT   08-DEC-1998 (first entry)
DE   3' fragment of prostate tumour specific gene; human; prostate cancer; detection;
KW   Prostate tumour specific gene; human; prostate cancer; detection;
     therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 39; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,

Query Match   19.7%;   Score 673.4; DB 1;   Length 789;
Best Local Similarity   94.2%;   Pred. No. 1.2e-98;
Matches 745; Conservative   0; Mismatches 40; Indels 6; Gaps

Qy 1341 CAGTCTATTTGGCCAGTGTGGCAGCTTCCCTGTGCCACATGCCTGTCCC 1400
         |||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1 CAGTCTATNTGGCCAGTGTGGCAGCTTCCCTGTGCCGGTGCCACATGCCTGTCCC 60

Qy 1401 ACAGTGTGCCGTGTGACAGCTTCAGCGCGCCCTCACCTTCTCAGCCCTGC 1460
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACAGTGTGCCGTGTGACAGCTTCAGCGCGCCCTCACCTTCTCAGCCCTGC 120

Qy 1461 AGATCCTGCCCTACACACTGGCCTCCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCA 1520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 AGATCCTGCCCTACACACTGGCCTCCCCTCTACCACCGGGAGAAGCAGGTGTTCCTGCCCA 180

Qy 1521 AATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGC 1580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AATACCGAGGGGACACTGGAGGTGCTAGCAGTGAGGACAGCCTGATGACCAGCTTCCTGC 240

Qy 1581 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCC 1640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGGACACGTGGGTGCTGGAGGCAGTGGCC 300

Qy 1641 TGCTCCCACCTCCACCCGCGCTCTGCGGGGCCTCTGCGGGTGGTTCGGGTACGTGTGG 1700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 TGCTCCCACCTCCACCCGCGCTCTGCGGGGCCTCTGCGGGTGGTTCGGGTACGTGTGG 360

Qy 1701 TGGTGGGTGAGCCCTACACACTGCCTGACCTCG 1760
         ||||||||||||||||||||||||||||||||
Db 361 TGGTGGGTGAGCCCTACACACTGCCTGACCTCG 420

Qy 1761 CCATCCTGGATAGTGCCTTCCTGCTGTTTATGGGGCTCCA 1820
         ||||||||||||||||||||||||||||||||||||||||
Db 421 CCATCCTGGATAGTG-CTTCCTGCTGTCCANGTGGCCCCATCCTGTTTATGGGGCTCCA 479

Qy 1821 TTGTCCAGCTCAGCCAGTGCTCTGCTGTCACTGCCTATATGGTGTCT 1880
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db 480 TTGTCCAGCTCAGCCAGTGTCTGCTGTCACTGCTCTATGGTGTCT 539

Qy 1881 TCGCCATTACTTTGCTACACAGGTAGTATTTGACAAGAGCGACTTGGCCAAATACTTCAG 1940
         |||||||||||||||||||||||||||||||||||||||||
Db 540 TC-CCATTTACTTTGCTACACAGGTANTATTTGACAAGAACGANTTGGCCAAATACTTCAG 598

Qy 1941 CGTAGAAAACTTCCAGCA--CATTGGGGTGGAGGGCCTGCCTCACTGGTCCCAGCTCG 1998
         ||| ||| ||||||||| |||||||||| |||||| |||||||||||||||||||
Db 599 CGTTAAAAAATTCCAGCAACATTGGGGTGGAAGGCCTGCCTCACTGGGT-CCAACTCCC 657

Qy 1999 CGCTCCTGTTAGCCCCATGGGGGCTGCCGGGGTCTGCCGCCCAGTTTCTGTTGCTGCCAAAGT 2058
         |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||

RESULT 6
V61144
ID V61144 standard; cDNA; 789 BP.
AC V61144;
DT 06-JAN-1999 (first entry)
DE 3' cDNA sequence of prostate tumour clone L1-12.

```
Db    658  CGCTCCTGTTAACCCCATGGGGCTGCCGGCTTGGCCGCCAATTTCTGTTGCTGCCAAANT  717
                ||||||||||||||||||||||||||||| ||||| ||||| ||| ||| |||
Qy   2059  AAGTGGCTCTCGCTGCTGCT-GCTGAGGTGCGTAGCTGCCAGCTGGGGGC  2117
                ||| |||||||||||||||| | ||||  |||| || |||
Db    718  NATGTGGCTCTCGCTGCTGCTGCTGAGTGCNTACNGCNCANCTNGGGGGG  777

Qy   2118  TGGGGCGTCCC  2128
                ||| || ||
Db    778  TNGGGNGTTCC  788
```

```
RESULT 7
V58488/c
ID   V58488 standard; cDNA; 772 BP.
AC   V58488;
DT   08-DEC-1998  (first entry)
DE   5' fragment of prostate tumour specific gene L1-12.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-805596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 39-40; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 772 BP;  221 A;  202 C;  170 G;  162 T;

Query Match    17.7%;  Score 603;  DB 1;  Length 772;
Best Local Similarity   91.8%;  Pred. No. 1.6e-87;
Matches  675;  Conservative  0;  Mismatches  49;  Indels  11;  Gaps  4;

Qy   1961  TGGGGTGGAGGGCCTGCCTCACTGGGTCCCCAGCTCCCCCGCTCCT----GTTAGCCCCA  2015
              ||||||||||||||||||| | |||||||||||||||||  ||||
Db    735  TNGGGGTGGNGGGCCTGCCTNAATTGGGTCCGNGTNCCCNGNTTCCTTAACCCCCN  676
Qy   2016  TGGGGCTGCCGGCTGGCCGCC---AGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCT-  2072
              | ||| |||||| ||||| ||    |||||| || ||| |||| ||
Db    675  TNGGGGTGCCGGGGTTGCCCCCCNAGTTTTTGTTGTTGCCAAAGTAAAGTGGTTTTTT  616
Qy   2072  GCTGCCACCCTGCTGCTGCTGAGTGCGTAGCTGCACAGCTGGGGCGTCCCTCT  2131
              ||||||||||||||||||| |||||| || |||||||||||| ||||||| |
Db    615  GCTGCCACCCTGCTGTTGAGTGGCTAGNTGCACAGNTGCTAGNTGCACAGCTGGGGNTGGGGGCGTCCTTTT  556
```

```
RESULT 8
V61145/c
ID   V61145 standard; cDNA; 772 BP.
AC   V61145;
DT   06-JAN-1999  (first entry)
DE   5' cDNA sequence of prostate tumour clone L1-12.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998;  U03492.
PR   09-FEB-1998; US-020955.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 39; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 772 BP;  221 A;  202 C;  170 G;  162 T;

Query Match    17.7%;  Score 603;  DB 1;  Length 772;
Best Local Similarity   91.8%;  Pred. No. 1.6e-87;
Matches  675;  Conservative  0;  Mismatches  49;  Indels  11;  Gaps  4;

Qy   1961  TGGGGTGGAGGGCCTGCCTCACTGGGTCCCCAGCTCCCCCGCTCCT----GTTAGCCCCA  2015
              ||||||||||||||||||| | |||||||||||||||||  ||||
Db    735  TNGGGGTGGNGGGCCTGCCTNAATTGGGTTCCANGTNCCCNGNTTCCTGTTAACCCCCN  676
Qy   2016  TGGGGCTGCCGGCTGGCCGCC---AGTTTCTGTTGCTGCCAAAGTAATGTGGCTCTCT-  2072
              | ||| |||||| ||||| ||    |||||| || ||| |||| ||
Db    675  TNGGGGTGCCGGGGTTGCCCCCCNAGTTTTTGTTGTTGCCAAAGTAAAGTGGTTTTTT  616
Qy   2072  GCTGCCACCCTGCTGCTGCTGAGTGCGTAGCTGCACAGCTGGGGCGTCCCTCT  2131
              ||||||||||||||||||| |||||| || ||||||||||| ||||||||
Db    615  GCTGCCACCCTGCTGTTGAGTGGCTAGNTGCACAGNTGCTAGNTGCACAGCTGGGGNTGGGGGNTGGGGGCGTCCTTTT  556
```

```
Db    375  TTTGGGAGCTGAATAAACTCAGTCACCTGGTTTCCCATCTNTAAGCCCCTTAACCTGCAG  316
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2372  CTTCGTTTAATGTAGCTCTTGCATGGGAAGAGTTCTAGGATGAAACACTCCTCATGGGATT  2431
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315  CTTCGTTTAATGTAGCTCTTGCATGGGAAGAGTTCTAGGATGAAACACTCCTCATGGGATT  256

Qy   2432  TGAACATATG--ACTTATTGTAGGGGAAGAGTCCTGAGGGGCAACACACAGAACCAGG  2489
               ||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db    255  TGAACATGAAAGTATTGTAGGGGAAGAGTCCTGAGGGGCAACACAAGAACCAGG  196

Qy   2490  TCCCCTCAGCCCACAGCACTGTCTTTTTGCTGATCCACCCCCCCTCTTACCTTTTATCAGG  2549
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195  TCCCCTCAGCCCACAGCACTGTCTTTTTGCTGATCCACCCCCCCTCTTACCTTTTATCAGG  136

Qy   2550  ATGTGGCCTGTTGGTCTTCGTTCGTTGCCATCACAGAGACACAGGCATTAAATATTAACT  2609
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135  ATGTGGCCTGTTGGTCTTCGTTCGTTGCCATCACAGAGACACAGGCATTAAATATTAACT  76

Qy   2610  TATTATTAACAAAGTAGAAGGAATCCATTGCTAGCTTTCTGTGTTGGTGTCTAATA  2669
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75   TATTATTAACAAAGTAGAAGGAATCCATTGCTAGCTTTCTGTGTTGGTGTCTAATA  16

Qy   2670  TTTGGGTAGGGTGGG  2684
               |||||||||||||||
Db    15   TTTGGGTAGGGTGGG  1
```

```
CC   differentiation activity, haematopoiesis regulating activity, tissue
CC   growth regulating activity, reproductive hormone regulating activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC   activity or other activities. The products can be used in forensic, gene
CC   therapy and chromosome mapping procedures. The sequences can also be used
CC   for obtaining corresponding promoter sequences. The nucleic acids
CC   encoding the signal peptide can be used for directing extracellular
CC   secretion of a polypeptide or the insertion of a polypeptide into a
CC   membrane, or importing a polypeptide into a cell.
SQ   Sequence  435 BP;    81 A;  126 C;  153 G;   69 T;
```

```
Query Match             11.9%;  Score 405.8;  DB 1;  Length 435;
Best Local Similarity   98.3%;  Pred. No. 2.3e-56;
Matches  414;  Conservative   0;  Mismatches   6;  Indels   1;  Gaps   1;
```

```
Qy      2 GGAACCAGCCTGCACGCGCTGGCTCCGGGTGACAGCCGGCGCGCTCGGCGCCAGGATCTGAG   61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     16 GGAACCAGCCTGCACGCGCTGGCTCCGGGTGACAGCCGGCGCGCTCGGCGCCAGGATCTGAG   75

Qy     62 TGATGAGACGTGTCCCACTGAGGTGCCCACAGCAGCAGGTGTTGAGCATGGGCTGAGA  121
          |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 TGAATGAGAAGTGTCCCACTGAGGTGCCCACACAGCAGCAGGTGTTGAGCATGGGCTGAGA  135

Qy    122 AGCTGGACCGGCACCAAAGGGCTGGCAGAAATGGCCGCCTGGCTGATTCCTAGGCAGTTG  181
          ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db    136 AGCTGGACCGGCACCAAAGGGCTGGCAGAAATDVGCGCCTGGCTGATTCCTAGGCAGTTG  195

Qy    182 GCGGCAGCAAGGAGGAGAGGCCCGAGCTTCTGGAGCAGAGCCCGAGACGGAGCAGTTCTGG  241
          |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GCRGCAGCAAGGAGGAGAGGCCCGAGCTTCTGGAGCAGAGCCCGAGACGGAGCAGTTCTGG  255

Qy    242 AGTGCCTGAACGGCCCCCTGAGCCTACCGCCGCCTGGCCCACCATGGTCCAGAGGCTGTGG  301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 AGTGCCTGAACGGCCCCCTGAGCCTACCGCCGCCTGGCCCACCATGGTCCAGAGGCTGTGG  315

Qy    302 GTGAGCCCCTGCTGCGGCACCGGAAAGCCCGAAAGCCCTCTTGCTGGTCAACCTGCTAACCTTT  361
          |||||||||||||||||||||||||||||||::||||||||::::|||||||||||||||
Db    316 GTGAGCCCCTGCTGCGGCACCGGAAAGCCCAGCTCCKKGCTGGKGAACCTGCTAACCTTT  375

Qy    362 GGCCTGGAGGTGTGTTGCGGCGAGCATCACCTATGTCCGCGCCTCTGCTGCTGGAAGTG  421
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    376 GGCCTGGAGGTGTGTTGGCGGCGCAGG-ATCACCTATGTGCCGCGCCTCTGCTGCTGGAAGTG  434

Qy    422 G 422
          |
Db    435 G 435
```

```
RESULT  10
V71173
ID   V71173 standard; cDNA; 342 BP.
AC   V71173;
DT   12-FEB-1999  (first entry)
DE   PS108 gene-specific clone 3520833.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9950567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
```

```
Qy   2132 CCTCTCTCCCAGTCTCCTAGGGCTGCCTGACTGGAGGCCTTCAGTCTG  2191
          | ||  ||||||||||||||| | |||||||||||||||| ||||||||
Db    555 CTTTTTCCCCAGTTTNTAGGGCTGCCTGACTGGAGNCCTTCCAAGGGGGTTCAGTCTG   496

Qy   2192 GACTTATACAGGAGGACCAGAAGGGCTTCCATGCACTCGAATGCGGGGACTCTGCAGGTGG  2251
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 GACTTATACAGGAGGACCAGAGAGGGTTCCATGCACTCGGAATGCGGGGACTCTGCAGGTGG   436

Qy   2252 ATTACCCAGGCTCAGGGTAACAGCTAGCCTCCTAGTTGAGACACCTAGAGAAGGGTT  2311
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 ATTACCCAGGCTCAGGGTAACAGCTAGCCTCCTAGTTGAGACACCTAGAGAAGGGTT   376

Qy   2312 TTTGGGAGCTGAATAAACTCAGTCACCTGGTTCCCATCTTCTTAAGCCCCTTAACCTGCAG  2371
          |||||||||||||||||||||||||||||||||||||||   |||||||||||||||||||
Db    375 TTTGGGAGCTGAATAAACTCAGTCACCTGGTTCCCATCTNTCAAGCCCCTTAACCTGCAG   316

Qy   2372 CTTCGTTTAATGTAGCTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCATGGGATT  2431
          |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 CTTCGTTTAAATGTAGCTCTTGCATGGGAGTTTCTAGGATGAAACACTCCTCCCATGGGATT   256

Qy   2432 TGAACATATG--ACTTATTTGTAGGGGAAGAGTCCTGAGGGGCAACACACAGAACCAGG  2489
          ||||||||||  | ||||||||||||||||||||||||||||||||||||||||||||||
Db    255 TGAACATATGAAAGTTATTTGTAGGGGAAGAGTCCTGAGGGGCAACACACAGAACCAGG  196

Qy   2490 TCCCCTCAGCCCCACAGCACTGTCTTTTTGCTGATCCACCCCCCTCTTATCCTTTATCAGG  2549
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 TCCCCTCAGCCCCACAGCACTGTCTTTTTGCTGATCCACCCCCCTCTTACCTTTATCAGG   136

Qy   2550 ATGTGGCCTGTTGGTCCTTCTGTTGCCATCACAGAGACACAGGCATTAAATATTTAAACT  2609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 ATGTGGCCTGTTGGTCCTTCTGTTGCCATCACAGAGACACAGGCATTAAATATTTAAACT   76

Qy   2610 TATTTATTTAACAAAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATA  2669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 TATTTATTTAACAAAGTAGAAGGGAATCCATTGCTAGCTTTTCTGTGTTGGTGTCTAATA   16

Qy   2670 TTTGGGTAGGGTGGG  2684
          |||||||||||||||
Db     15 TTTGGGTAGGGTGGG  1
```

```
RESULT   9
X40506
ID   X40506 standard; cDNA; 435 BP.
AC   X40506;
DT   18-JUN-1999  (first entry)
DE   Human secreted protein 5' EST SEQ ID No: 106.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS   Homo sapiens.
PN   WO9905550-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1232.
PR   01-AUG-1997; US-905144.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153780/13.
DR   P-PSDB; Y11784.
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
PS   Claim 1: Page 224; 675pp; English.
CC   X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins expressed in prostate, and encode the proteins given in
CC   Y11716 to Y11993 respectively. The proteins given represent the signal
CC   peptide and an N-terminal fragment of a secreted protein. The nucleic
CC   acid sequences can be used for producing secreted human gene products.
CC   They can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell proliferation and
```

```
CC   Nucleotide sequences v71166-79 represent overlapping clones comprising
CC   the sequence v71180 and the consensus sequence v71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 342 BP;  84 A;  88 C;  81 G;  88 T;

Query Match             9.4%;  Score 319.8;  DB 1;  Length 342;
Best Local Similarity  98.5%;
Matches  333;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    1;

Qy   2207 GCCAGAAGGGCTCCATGCACTGCACTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAG 2266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCAGAAGGGCTCCATGCACTGCACTGGAATGCGGGGACTCTGCAGGTGGATTACCCAGGCTCAG   60

Qy   2267 GGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGGTTTTGGGAGCTGAATA 2326
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGTTAACAGCTAGCCTCCTAGTTGAGACACACCTAGAGAAGGGGTTTTGGGAGCTGAATA  120

Qy   2327 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAG 2386
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AACTCAGTCACCTGGTTTCCCATCTCTAAGCCCCTTAACCTGCAGCTTCGTTTAATGTAG  180

Qy   2387 CTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCCATGGGATTGAACATATG--ACT 2444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    181 CTCTTGCATGGGAGTTCTAGGATGAAACACTCCTCCCATGGGATTGAACATATGAAAGT  240

Qy   2445 TATTTGTAGGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAGGTCCCCTCAGCCCACA 2504
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TATTTGTAGGGGAAGAGTCCTGAGGGGCAACACACAAGAACCAGGTCCCCTCAGCCCACA  300

Qy   2505 GCACTGTCTTTTTGCTGATCCACCCCCCTTACCTTT 2542
           ||||||||||||||||||||| ||||||||| |||||
Db    301 GCACTGTCTTTTTGCTGATCCANCCCCCTCTTACTTTT  338

RESULT  11
V71177
ID   V71177 standard; cDNA; 294 BP.
AC   V71177;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 3705332.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO985067-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences v71166-79 represent overlapping clones comprising
CC   the sequence v71180 and the consensus sequence v71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 294 BP;  87 C;  60 G;  74 T;

Query Match             8.5%;  Score 291;  DB 1;  Length 294;
Best Local Similarity  99.0%;  Pred. No. 3.1e-38;
Matches  291;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   2974 AAGGCACTGCCCAAAATTCCCCTACCCCCAACTTCCCCTACCCCCAACTTCCCCACC 3033
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      1 AAGGCACTGCCCAAAATNNCCCTACCCCCAACTTCCCCTACCCCCAACTTCCCCACC   60

Qy   3034 AGCTTCCACAACCCTGTTTGGAGCTACTGCGAGGACCAGAAGCACAAAGTGCGGTTTCCCAA 3093
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGCTCCCACAACCCTGTTTGGAGCTACTGCGAGGACCAGAAGCACAAAGTGCGGTTTCCCAA  120

Qy   3094 GCCTTTGTCCATCTCAGCCCCCAGTATATCTGCTTGGGAATCTCACACAGAAACT 3153
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCCTTTGTCCATCTCAGCCCCCAGTATATCTGCTTGGGAATCTCACACAGAAACT  180

Qy   3154 CAGGAGCACCCCCTGCCTGAGCTAAGGAGGTCTTATCTCCAGGGGGGTTTAAGTGCC 3213
           |||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    181 CAGGAGCACCCCCTGCNTGAGCTAAGGAGGTCTTATCTCCAGGGGGGTTTAAGTGCC  240

Qy   3214 GTTTGCAATAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAA 3267
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTTTGCAATAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAA  294

RESULT  12
V71175
ID   V71175 standard; cDNA; 288 BP.
AC   V71175;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 3497504.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO985067-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences v71166-79 represent overlapping clones comprising
CC   the sequence v71180 and the consensus sequence v71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 288 BP;  70 A;  66 C;  61 G;  90 T;
```

Query Match   8.4%;   Score 287; DB 1; Length 288;
Best Local Similarity   99.7%   Pred. No. 1.3e-37;
Matches 287;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

Qy 2532 CTCTTACCTTTATCAGGATGTGGCCGTGTTGGTCCTTCTGTTGCCATCACAGAGACACAG 2591
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CTCTTACCTTTATCAGGATGTGGCCGTGTTGGTCCTTCTGTTGCCATCACAGAGACACAG 60

Qy 2592 GCATTAAATATTTAACTTATTATTATTAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 2651
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCATTAAATATTTAACTTATTATTATTAACAAAGTAGAAGGAATCCATTGCTAGCTTTT 120

Qy 2652 CTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCCAACAAATCAGGTCCCCTGAG 2711
        ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db 121 CTGTGTTGGTGTCTAATATTTGGGTANGGTGGGGATCCCCAACAATCAGGTCCCCTGAG 180

Qy 2712 CTGTGTTGGTGTCTAATATTTGGGTAGGGTGGGGATCCCCAACAATCAGGTCCCCTGAG 2771
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ATAGCTGGTCATTGGGCTGATCATTGCCAGAATCTTCTTCTCCTGGGGTCTGGCCCCCCA 240

Qy 2772 AAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAATGATAA 2819
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 AAATGCCTAACCCAGGACCTTGGAAATTCTACTCATCCCAAATGATAA 288

RESULT 13
V71176
ID   V71176 standard; cDNA; 272 BP.
AC   V71176;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 3964174.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 272 BP; 58 A; 96 C; 49 G; 68 T;

Query Match   7.9%;   Score 270; DB 1; Length 272;
Best Local Similarity   99.6%   Pred. No. 6.4e-35;
Matches 270;   Conservative 0;   Mismatches 0;   Indels 1;   Gaps 0;

Qy 2795 AAATTCTACTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAA 2854
        |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   1 AAATTCTACTCATCCCAAATGATAATTCCAAATGCTGTTACCCAAGGTTAGGGTGTTGAA 60

Qy 2855 GGAAGGTAGAGGGTGGGGCCTTCAGGTCTCAAGCGGCTTCCTCAACACCCTTCCTCCTTG 2914
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGAAGGTAGAGGGTGGGGCCTTCAGGTCTCAAGCGGCTTCCTCAACACCCTTCCTCCTTG 120

Qy 2915 GCCCAGCCTGGTTCCCCCACTTCCACTTCCACTCTCTCTAGGACTGGGCTGATGA 2974
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GCCCAGCCTGGTTCCCCCACTTCCACTTCCACTCTCTCTAGGACTGGGCTGATGA 180

Qy 2975 AGGCACTGCCCAAAATTTCCCCTACCCCCAACTTTCCCCTACCCCCACCA 3034
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AGGCACTGCCCAAAATTTCCCCTACCCCCAACTTTCCCCTACCCCCACCA 240

Qy 3035 GCTCCACAACCCTGTTTGGAGCTACTGCAGG 3065
        |||||||||||||||||||||||||||||||
Db 241 GCTCCACAACCCTGTTTGGAGCTACTGCAGG 271

RESULT 14
V71174
ID   V71174 standard; cDNA; 265 BP.
AC   V71174;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 2188949.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI; 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence 265 BP; 64 A; 64 C; 61 G; 76 T;

Query Match   7.8%;   Score 265; DB 1; Length 265;
Best Local Similarity   100.0%   Pred. No. 4e-34;
Matches 265;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy 2453 GGGGAGAGTCCTGAGGGGCAACACACAAGAACCAGGTCCCCTCAGCCCAGCCACTGTC 2512
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGGAGAGTCCTGAGGGGCAACACACAAGAACCAGGTCCCCTCAGCCCAGCCACTGTC 60

Qy 2513 TTTTTGCTGATCCACCCCCCTTACCTTTTATCAGGATGTGGCCTTGTTGGTCCTTCTGT 2572
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTTTTGCTGATCCACCCCCCTTACCTTTTATCAGGATGTGGCCTTGTTGGTCCTTCTGT 120

Qy 2573 TGCCATCACAGAGACACAGGCATTTAAATATTTAACTTATTTATTATTTAACAAAGTAGAAGG 2632
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 TGCCATCACAGAGACACAGGCATTTAAATATTTAACTTATTTATTATTTAACTTAACAAAGTAGAAGG 180

Search completed: September 25, 1999, 09:55:31
Job time: 7982 sec

```
Qy   2633 GAATCCATTGCTAGCTTTCTGTGTTGGTGTGCTAATATTTGGGTAGGTGGGGATCCCC 2692
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAATCCATTGCTAGCTTTCTGTGTTGGTGTGCTAATATTTGGGTAGGTGGGGATCCCC 240

Qy   2693 AACAATCAGGTCCCCTGAGATAGCT 2717
          |||||||||||||||||||||||||
Db    241 AACAATCAGGTCCCCTGAGATAGCT 265


RESULT  15
V71166
ID   V71166 standard; cDNA; 258 BP.
AC   V71166;
DT   12-FEB-1999 (first entry)
DE   PS108 gene-specific clone 1864683.
KW   PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW   prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW   drug screening; gene therapy; ss.
OS   Homo sapiens.
PN   WO9850567-A1.
PD   12-NOV-1998.
PF   01-MAY-1998; U08930.
PR   02-MAY-1997; US-850713.
PA   (ABBO ) ABBOTT LAB.
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI   Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD,
PI   Roberts-Rapp L, Russell JC, Stroupe SD;
DR   WPI: 99-034731/03.
PT   New isolated prostate-specific polynucleotides - used to develop
PT   products for the diagnosis and treatment of prostate diseases, e.g.
PT   benign hyperplasia, prostatic or prostate cancer
PS   Claim 1; Fig 1A-E; 122pp; English.
CC   Nucleotide sequences V71166-79 represent overlapping clones comprising
CC   the sequence V71180 and the consensus sequence V71181. The sequences
CC   are PS108 gene-specific. They are used in the method of the invention.
CC   The specification describes a method for detecting the presence of a
CC   target PS108 polynucleotide in a test sample. The method comprises
CC   contacting the test sample with at least 1 PS108-specific polynucleotide
CC   or complement, and detecting the presence of the target PS108
CC   polynucleotide. The products can be used for detecting, diagnosing,
CC   staging, monitoring, prognosticating, in vivo imaging, preventing or
CC   treating, or determining predisposition to diseases or conditions of
CC   the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC   prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC   products can be used in drug screening and gene therapy.
SQ   Sequence  258 BP;   35 A;   78 C;   88 G;   57 T;

Query Match             7.5%;  Score 256.4;  DB 1;  Length 258;
Best Local Similarity  99.6%;  Pred. No. 9e-33;
Matches  257;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1177 GGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGCACCGAGGCCCGGAGACACTATGA 1236
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGGCTGTACCAGGGCGTGCCCAGAGCTGAGCCGGCACCGAGGCCCGGAGACACTATGA 60

Qy   1237 TGAAGGCGTTCGGATGGGCAGCCTGGGCGTGTTCCTGCAGTGCGCCATCTCCTGGTCTT 1296
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGAAGGCGTTCGGATGGGCAGCCTGGGCGTGTTCCTGCAGTGCGCCATCTCCTGGTCTT 120

Qy   1297 CTCTCTGGTCATGGACCGGCTGGTGCCGTCGCCGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 1356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCTCTGGTCATGGACCGGCTGGTGCCGTCGCCGGATTCGGCACTCGAGCAGTCTATTTGGCCAG 180

Qy   1357 TGTGGCAGCTTTCCCTGTGGCTGCCCGTGCCACATGCCTGTCCCCACAGTGTGGCCGTGGT 1416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTGGCAGCTTTCCCTGTGGCTGCCCGTGCCACATGCCTGTCCCCACAGTGTGGCCGTGGT 240

Qy   1417 GACAGCTTCAGCCGCCCT 1434
          |||||||||||||| |
Db    241 GACAGCTTCAGCCGCCTT 258
```

us-09-030-606-110_1.rng

```
                   GenCore version 4.5
          Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   September 25, 1999, 06:32:43 ; Search time 165.01 Seconds
                              (without alignments)
                              1894.534 Million cell updates/sec

Title:           US-09-030-606-110
Perfect score:   3410
Sequence:        1 GGGAACCAGCCTGCACGCGC.........AAAAAATAAAAAAAAAA 3410

Scoring table:   IDENTITY_NUC

Searched:        176461 seqs, 45838279 residues

Database :       Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	91.4	2.7	1066	3	US-08-157-101A-4	Sequence 4, Appli
2	90.8	2.7	6671	2	US-08-280-443A-1	Sequence 1, Appli
3	90.8	2.7	6671	2	US-08-457-459-1	Sequence 1, Appli
4	90.8	2.7	6671	2	US-08-555-678-1	Sequence 1, Appli
5	90.8	2.7	6671	5	PCT-US95-02275-1	Sequence 1, Appli
6	89.4	2.6	144	2	US-08-702-344-26	Sequence 26, Appli
7	89	2.6	1641	1	US-08-300-903A-8	Sequence 8, Appli
8	88.2	2.6	140	1	US-08-628-417-5	Sequence 5, Appli
9	88.2	2.6	240	1	US-08-628-417-6	Sequence 6, Appli
10	88.2	2.6	1602	3	US-08-330-950-3	Sequence 3, Appli
11	87.8	2.6	1493	1	US-08-330-820-24	Sequence 24, Appli
12	87.8	2.5	1493	1	US-08-593-535-24	Sequence 24, Appli
13	86.6	2.5	1817	1	US-08-473-981A-5	Sequence 5, Appli
14	85	2.5	903	5	PCT-US95-06406A-21	Sequence 21, Appli
15	84	2.5	2082	4	US-08-785-310A-2	Sequence 2, Appli
16	83.8	2.5	635	3	US-08-455-633A-35	Sequence 35, Appli
17	83.8	2.5	635	3	US-08-416-335-5	Sequence 5, Appli
18	83.8	2.5	635	5	PCT-US94-05354-35	Sequence 35, Appli
19	83.4	2.5	630	1	US-08-185-414E-1	Sequence 1, Appli
20	83	2.4	9589	1	US-07-925-695-1	Sequence 1, Appli
c 21	83	2.4	9589	1	US-07-925-695-2	Sequence 2, Appli
c 22	83	2.4	98	1	US-08-088-658-42	Sequence 42, Appli
23	83	2.4	117	2	US-08-702-344-3	Sequence 3, Appli
24	81.2	2.4	1332	4	US-09-057-762-1	Sequence 1, Appli
25	81	2.4	1172	1	US-07-945-288-9	Sequence 9, Appli
26	81	2.4	1172	1	US-08-462-83L-9	Sequence 9, Appli
27	81	2.4	1172	3	US-08-461-809-9	Sequence 9, Appli
28	81	2.4	1172	1	US-08-461-441-9	Sequence 9, Appli
29	81	2.4	1172	5	PCT-US93-08518-9	Sequence 9, Appli
30	79.6	2.3	270	4	US-08-520-678A-30	Sequence 30, Appl
c 31	79.2	2.3	266	4	US-08-520-678A-29	Sequence 29, Appl
c 32	78.2	2.3	356	4	US-08-520-678A-22	Sequence 22, Appl
33	77.8	2.3	1147	3	US-08-665-716-1	Sequence 1, Appli
c 34	77.8	2.3	84	3	US-08-664-596B-3	Sequence 3, Appli
35	77.8	2.3	84	3	US-08-738-367-3	Sequence 3, Appli
36	77.6	2.3	1921	4	US-08-557-128-11	Sequence 11, Appl
37	76.8	2.3	11517	3	US-07-920-281C-1	Sequence 1, Appli
38	76.6	2.2	2010	1	US-07-864-475A-4	Sequence 4, Appli
39	76.6	2.2	2010	4	US-08-468-249A-4	Sequence 4, Appli
40	76.4	2.2	1023	1	US-08-252-966B-16	Sequence 16, Appl
41	76.4	2.2	5173	4	US-08-242-677-1	Sequence 1, Appli
42	76.4	2.2	1325	4	US-08-464-517-1	Sequence 1, Appli
43	76.4	2.2	1325	5	PCT-US93-05000-1	Sequence 1, Appli
c 44	76.2	2.2	5852	1	US-07-867-106-2	Sequence 2, Appli
c 45	76.2	2.2	1325	3	US-08-306-691B-51	Sequence 51, Appl

```
                              ALIGNMENTS

RESULT    1
US-08-157-101A-4
; Sequence 4, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
;   APPLICANT: KURIHARA, TATSUYA
;   APPLICANT: MATSUKURA, SHIGEKAZU
;   APPLICANT: TSURUOKA, NOBUO
;   APPLICANT: ARIMA, KENJI
;   APPLICANT: NISHIHARA, TATSURO
;   TITLE OF INVENTION: ANTI-HBs ANTIBODY GENES AND EXPRESSION
;   TITLE OF INVENTION: PLASMIDS THEREFOR
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: PILLSBURY, MADISON & SUTRO
;     STREET: 1100 NEW YORK AVENUE, N.W.
;     CITY: WASHINGTON
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/157,101A
;     FILING DATE: 05-APR-1994
;     CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;     NAME: TITUS, MARLANA K
;     REGISTRATION NUMBER: 35843
;     REFERENCE/DOCKET NUMBER: 9437/204199
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-861-3711
;     TELEFAX: 202-822-0944
;     TELEX: 6714627 CUCH
; INFORMATION FOR SEQ ID NO:    4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1066 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-157-101A-4

  Query Match             2.7%;  Score 91.4;  DB 3;  Length 1066;
  Best Local Similarity  81.9%;  Pred. No. 7.2e-10;
  Matches  104; Conservative  1; Mismatches  22; Indels  0; Gaps  0;

Qy   3284 TAATGTTTATGTGACAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAA 3343
               ||||||| |||||| |||  || ||  |||||||||||||||||||||||
Db    898 TAATGTTGGAGGAATGAATAATAAGTGAATCTTTGCAAAAAAAAAAAAAA 957

Qy   3344 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAATAAAA 3403
               ||||||||||||||| |||||||||||||||||||||||||| ||||
Db    958 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1017

Qy   3404 AAAAAA 3410
```

Db 1018 AAAAAA 1024
          ||||||

RESULT 2
US-08-280-443-1
; Sequence 1, Application US/08280443
; Patent No. 5643778
; GENERAL INFORMATION:
;   APPLICANT: Nishikura, Kazuko
;   TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
;   TITLE OF INVENTION: Thereof
;   NUMBER OF SEQUENCES: 39
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Howson and Howson
;   STREET: Spring House Corporate Cntr, P.O. Box 457
;   CITY: Spring House
;   STATE: Pennsylvania
;   COUNTRY: USA
;   ZIP: 19477
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/280,443
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/197,794
;   FILING DATE: 17-FEB-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Bak, Mary E.
;   REGISTRATION NUMBER: 31,215
;   REFERENCE/DOCKET NUMBER: WST49AUSA
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-540-9206
;   TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6671 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 155..3832
US-08-280-443-1

Query Match   2.7%; Score 90.8; DB 2; Length 6671;
Best Local Similarity 76.4%; Pred. No. 1.7e-09;
  Matches 110; Conservative   1; Mismatches   33; Indels   0; Gaps   0;

Qy   3267 AGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
          ||||| |||||| ||| |||| | ||||||| ||
Db   6522 AGTGACTTAACAATATACATTCCTCATAAAATCAAGAATCTGAAAAAAAAA 6581

Qy   3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
          ||||||||||||||||||||||||||||||||||||||:||||||||||||
Db   6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAA 6641

Qy   3387 AAAAAAAAAATAAAAAAAAAAA 3410
          |||||||||| ||||||||||
Db   6642 AAAAAAAAAAAAAAAAAAAAAA 6665

RESULT 3
US-08-457-459-1
; Sequence 1, Application US/08457459
; Patent No. 5677428

GENERAL INFORMATION:
  APPLICANT: Nishikura, Kazuko
  TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
  TITLE OF INVENTION: Thereof
  NUMBER OF SEQUENCES: 39
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Howson and Howson
    STREET: Spring House Corporate Cntr, P.O. Box 457
    CITY: Spring House
    STATE: Pennsylvania
    COUNTRY: USA
    ZIP: 19477
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US 08/280,443
    FILING DATE: 25-JUL-1994
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/197,794
    FILING DATE: 17-FEB-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: Bak, Mary E.
    REGISTRATION NUMBER: 31,215
    REFERENCE/DOCKET NUMBER: WST49CUSA
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 215-540-9206
    TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 6671 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: unknown
  MOLECULE TYPE: cDNA
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 155..3832
US-08-457-459-1

Query Match   2.7%; Score 90.8; DB 2; Length 6671;
Best Local Similarity 76.4%; Pred. No. 1.7e-09;
  Matches 110; Conservative   1; Mismatches   33; Indels   0; Gaps   0;

Qy   3267 AGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
          ||||| |||||| ||| |||| | ||||||| ||
Db   6522 AGTGACTTAACAATATACATTCCTCATAAAATCAAGAATCTGAAAAAAAAA 6581

Qy   3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
          ||||||||||||||||||||||||||||||||||||||:||||||||||||
Db   6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAA 6641

Qy   3387 AAAAAAAAAATAAAAAAAAAAA 3410
          |||||||||| ||||||||||
Db   6642 AAAAAAAAAAAAAAAAAAAAAA 6665

RESULT 4
US-08-555-678-1
; Sequence 1, Application US/08555678
; Patent No. 5763174
; GENERAL INFORMATION:
;   APPLICANT: Nishikura, Kazuko
;   TITLE OF INVENTION: RNA Editing Enzyme and Methods
;   TITLE OF INVENTION: of Use Thereof
;   NUMBER OF SEQUENCES: 67

```
; CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Howson and Howson
;    STREET: Spring House Corporate Cntr, P.O. Box 457
;    CITY: Spring House
;    STATE: Pennsylvania
;    COUNTRY: USA
;    ZIP: 19477
; COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/555,678
;    FILING DATE:
;    CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/197,794
;    FILING DATE: 17-FEB-1994
; PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/280,443
;    FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/457,459
;    FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
;    NAME: Bak, Mary E.
;    REGISTRATION NUMBER: 31,215
;    REFERENCE/DOCKET NUMBER: WST49DUSA
; TELECOMMUNICATION INFORMATION:
;    TELEPHONE: 215-540-9206
;    TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 6671 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: double
;    TOPOLOGY: unknown
;    MOLECULE TYPE: cDNA
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 155..3832
US-08-555-678-1

Query Match              2.7%;  Score 90.8;  DB 3;  Length 6671;
Best Local Similarity   76.4%;  Pred. No. 1.7e-09;
Matches  110;  Conservative    1;  Mismatches   33;  Indels    0;  Gaps    0;

Qy   3267 AGTGAGCAATCAGAGTATATGTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
          |||||  || || ||| ||| ||   || ||   ||||    ||| ||  |  || |
Db   6522 AGTGACTTAACAATATACATTCCTCATAAATAAATAAAAAAAAACAAGAATCTGAAAAAAAA 6581

Qy   3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
          ||||||||||||||||||||||||||||||||||::|||||||||||||||||||||
Db   6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6641

Qy   3387 AAAAAAAAAAATAAAAAAAAAA 3410
          ||||||||||| ||||||||||
Db   6642 AAAAAAAAAAAAAAAAAAAAAA 6665

RESULT 5
PCT-US95-02275-1
; Sequence 1, Application PCT/TUS9502275
; GENERAL INFORMATION:
;    APPLICANT: Wistar Institute of Anatomy & Biology
;    TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
;    TITLE OF INVENTION: Thereof
;    NUMBER OF SEQUENCES: 39
;    CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Howson and Howson
;    STREET: Spring House Corporate Cntr, P.O. Box 457
```

```
;    CITY: Spring House
;    STATE: Pennsylvania
;    COUNTRY: USA
;    ZIP: 19477
; COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: PCT/US95/02275
;    FILING DATE:
;    CLASSIFICATION:
; PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/280,443
;    FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 08/197,794
;    FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
;    NAME: Bak, Mary E.
;    REGISTRATION NUMBER: 31,215
;    REFERENCE/DOCKET NUMBER: WST49BPCT
; TELECOMMUNICATION INFORMATION:
;    TELEPHONE: 215-540-9206
;    TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 6671 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: double
;    TOPOLOGY: unknown
;    MOLECULE TYPE: cDNA
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 155..3832
PCT-US95-02275-1

Query Match              2.7%;  Score 90.8;  DB 5;  Length 6671;
Best Local Similarity   76.4%;  Pred. No. 1.7e-09;
Matches  110;  Conservative    1;  Mismatches   33;  Indels    0;  Gaps    0;

Qy   3267 AGTGAGCAATCAGAGTATATGTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTA 3326
          |||||  || || ||| ||| ||   || ||   ||||    ||| ||  |  || |
Db   6522 AGTGACTTAACAATATACATTCCTCATAAATAAATAAAAAAAAACAAGAATCTGAAAAAAAA 6581

Qy   3327 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3386
          ||||||||||||||||||||||||||||||||||::|||||||||||||||||||||
Db   6582 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6641

Qy   3387 AAAAAAAAAAATAAAAAAAAAA 3410
          ||||||||||| ||||||||||
Db   6642 AAAAAAAAAAAAAAAAAAAAAA 6665

RESULT 6
US-08-702-344-26
; Sequence 26, Application US/08702344
; Patent No. 5723315
; GENERAL INFORMATION:
;    APPLICANT: Jacobs, Kenneth
;    APPLICANT: McCoy, John
;    APPLICANT: LaValle, Edward
;    APPLICANT: Racie, Lisa
;    APPLICANT: Merberg, David
;    APPLICANT: Treacy, Maurice
;    APPLICANT: Spaulding, Vikki
;    TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;    TITLE OF INVENTION: ENCODING THEM
;    NUMBER OF SEQUENCES: 37
;    CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Genetics Institute, Inc.
```

```
                STREET: 87 CambridgePark Drive
                CITY: Cambridge
                STATE: Massachusetts
                COUNTRY: U.S.A.
                ZIP: 02140
        COMPUTER READABLE FORM:
                MEDIUM TYPE: Floppy disk
                COMPUTER: IBM PC compatible
                OPERATING SYSTEM: PC-DOS/MS-DOS
                SOFTWARE: PatentIn Release #1.0, Version #1.30
        CURRENT APPLICATION DATA:
                APPLICATION NUMBER: US/08/702,344
                FILING DATE:
                CLASSIFICATION: 536
        ATTORNEY/AGENT INFORMATION:
                NAME: Brown, Scott A.
                REGISTRATION NUMBER: 32,724
        TELECOMMUNICATION INFORMATION:
                TELEPHONE: (617) 498-8224
                TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 26:
        SEQUENCE CHARACTERISTICS:
                LENGTH: 144 base pairs
                TYPE: nucleic acid
                STRANDEDNESS: double
                TOPOLOGY: linear
        MOLECULE TYPE: cDNA
US-08-702-344-26

  Query Match          2.6%;    Score 89.4; DB 2; Length 144;
  Best Local Similarity  85.0%;    Pred. No. 9.1e-10;  Indels  0;  Gaps  0;
  Matches 96;  Conservative  1;  Mismatches 16;

Qy  3298 ACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3357
         | ||| ||| || ||  ||   |||||||||||||||||||||||||||||||||
Db     1 AAAAAAAAAANNTNNTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy  3358 AAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAA 3410
         ||||||||||:||||||||||||||||||||| |||||||||
Db    61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 113

RESULT  7
US-08-300-903A-8
; Sequence 8, Application US/08300903A
; Patent No. 5591630
        GENERAL INFORMATION:
                APPLICANT: Anderson, Dirk M
                APPLICANT: Giri, Judith G
                TITLE OF INVENTION: Interleukin-15 Receptors
                NUMBER OF SEQUENCES: 15
                CORRESPONDENCE ADDRESS:
                ADDRESSEE: Immunex Corporation
                STREET: 51 University Street
                CITY: Seattle
                STATE: Washington
                COUNTRY: USA
                ZIP: 98101
        COMPUTER READABLE FORM:
                MEDIUM TYPE: Floppy disk
                COMPUTER: Apple Macintosh
                OPERATING SYSTEM: Apple Operating System 7.1
                SOFTWARE: Microsoft Word for Apple, Version 5.1a
        CURRENT APPLICATION DATA:
                APPLICATION NUMBER: US/08/300,903A
                FILING DATE: 06-SEPTEMBER-1994
                CLASSIFICATION: 435
        PRIOR APPLICATION DATA:
                APPLICATION NUMBER: USSN 08/236,919
                FILING DATE: 06-MAY-1994
                CLASSIFICATION: 435
        ATTORNEY/AGENT INFORMATION:
```

```
                NAME: Perkins, Patricia Anne
                REGISTRATION NUMBER: 34,695
                REFERENCE/DOCKET NUMBER: 2822-A
        TELECOMMUNICATION INFORMATION:
                TELEPHONE: 206-587-0430
                TELEFAX: 206-233-0644
INFORMATION FOR SEQ ID NO: 8:
        SEQUENCE CHARACTERISTICS:
                LENGTH: 1641 base pairs
                TYPE: nucleic acid
                STRANDEDNESS: single
                TOPOLOGY: linear
        MOLECULE TYPE: cDNA
        HYPOTHETICAL: NO
        ANTI-SENSE: NO
        FEATURE:
                NAME/KEY: CDS
                LOCATION: 3..839
US-08-300-903A-8

  Query Match          2.6%;    Score 89; DB 1; Length 1641;
  Best Local Similarity  85.2%;    Pred. No. 2.5e-09;  Indels  0;  Gaps  0;
  Matches 98;  Conservative  1;  Mismatches 16;

Qy  3296 TGACAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3355
         | | ||| |||| | ||| |  | ||| |   |||||||||||||||||||||||||||
Db  1525 TAAACAAAATAAAACATCTATTTCAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1584

Qy  3356 AAAAAAAAAAAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
         ||||||||||||:|||||||||||||||||||||||||||||| ||||||||||
Db  1585 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1639

RESULT  8
US-08-628-417-5
; Sequence 5, Application US/08628417
; Patent No. 5627054
        GENERAL INFORMATION:
                APPLICANT: GILLESPIE, DAVID
                TITLE OF INVENTION: COMPETITOR PRIMER ASYMETRIC
                TITLE OF INVENTION: POLYMERASE CHAIN REACTION
                NUMBER OF SEQUENCES: 7
                CORRESPONDENCE ADDRESS:
                ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
                ADDRESSEE: DEFENSE COMMAND
                STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
                CITY: ABERDEEN PROVING GROUND
                STATE: MARYLAND
                COUNTRY: USA
                ZIP: 21010-5423
        COMPUTER READABLE FORM:
                MEDIUM TYPE: Floppy disk
                COMPUTER: IBM PC compatible
                OPERATING SYSTEM: PC-DOS/MS-DOS
                SOFTWARE: PatentIn Release #1.0, Version #1.25
        CURRENT APPLICATION DATA:
                APPLICATION NUMBER: US/08/628,417
                FILING DATE:
                CLASSIFICATION: 435
        ATTORNEY/AGENT INFORMATION:
                NAME: BIFFONI, ULYSSES J
                REGISTRATION NUMBER: 39,908
                REFERENCE/DOCKET NUMBER: DAM 398-94
        TELECOMMUNICATION INFORMATION:
                TELEPHONE: 410-671-1158
                TELEFAX: 410-671-2534
INFORMATION FOR SEQ ID NO: 5:
        SEQUENCE CHARACTERISTICS:
                LENGTH: 140 bases
                TYPE: nucleic acid
                STRANDEDNESS: single
                TOPOLOGY: linear
```

; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-628-417-5

Query Match   2.6%;   Score 88.2;   DB 1;   Length 140;
Best Local Similarity   80.3%;   Pred. No. 1.6e-09;
Matches 102; Conservative   1; Mismatches 24; Indels 0; Gaps 0;

```
Qy 3284 TAATGTTTATGGTGACAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAA 3343
         || || || |||| || || |||| || | |||| ||||||||||||||||
Db    1 TAGAGATAAAGACAATAAAACTTTAGAAATAATTTACTAAAAAAAAAAAAAAAAA 60

Qy 3344 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAA 3403
         |||||||||||||||||||||||||||||||::||||||||||||||||||||||
Db   61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy 3404 AAAAAAA 3410
         |||||||
Db  121 AAAAAAA 127
```

RESULT 9
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; TITLE OF INVENTION: POLYMERASE CHAIN REACTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; ADDRESSEE: DEFENSE COMMAND
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,417
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BIFFONI, ULYSSES J
; REGISTRATION NUMBER: 39,908
; REFERENCE/DOCKET NUMBER: DAM 398-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 410-671-1158
; TELEFAX: 410-671-2534
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: oligodeoxynucleotide
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-628-417-6

Query Match   2.6%;   Score 88.2;   DB 1;   Length 240;
Best Local Similarity   80.3%;   Pred. No. 1.9e-09;
Matches 102; Conservative   1; Mismatches 24; Indels 0; Gaps 0;

```
Qy 3284 TAATGTTTATGGTGACAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAA 3343
         || || || |||| || || |||| || | |||| ||||||||||||||||
Db    1 TAGAGATAAAGACAATAAAACTTTAGAAATAATTTACTAAAAAAAAAAAAAAAAA 60

Qy 3344 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAA 3403
         |||||||||||||||||||||||||||||||::||||||||||||||||||||||
Db   61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 120

Qy 3404 AAAAAAA 3410
         |||||||
Db  121 AAAAAAA 127
```

RESULT 10
US-08-530-950-3
; Sequence 3, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-530-950-3

Query Match   2.6%;   Score 88.2;   DB 3;   Length 1602;
Best Local Similarity   86.5%;   Pred. No. 3.5e-09;
Matches 96; Conservative   1; Mismatches 14; Indels 0; Gaps 0;

```
Qy 3300 AAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3359
         || ||| || |||| | ||| ||| ||||||||||||||||||||||||||||||
Db 1490 AATATTTAATGATGTGTCATATGAGTCCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 1549

Qy 3360 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAA 3410
         ||||||||||||||||||||||||||||||||||||||| |||||||||||
Db 1550 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1600
```

RESULT 11
US-08-340-820-24

; Sequence 24, Application US/08340820
; Patent No. 5512460
; ,, GENERAL INFORMATION:
; ,, APPLICANT: NARUO, Ken-ichi
; ,, APPLICANT: SEKO, Chisako
; ,, APPLICANT: KUROKAWA, Tsutomu
; ,, APPLICANT: KONDO, Tatsuya
; ,, TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
; ,, TITLE OF INVENTION: PRODUCTION
; ,, NUMBER OF SEQUENCES: 27
; ,, CORRESPONDENCE ADDRESS:
; ,, ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ,, ADDRESSEE: CUSHMAN
; ,, STREET: 130 Water Street
; ,, CITY: Boston
; ,, STATE: Massachusetts
; ,, COUNTRY: US
; ,, ZIP: 02109
; ,, COMPUTER READABLE FORM:
; ,, MEDIUM TYPE: Floppy disk
; ,, COMPUTER: IBM PC compatible
; ,, OPERATING SYSTEM: PC-DOS/MS-DOS
; ,, SOFTWARE: PatentIn Release #1.0, Version #1.25
; ,, CURRENT APPLICATION DATA:
; ,, APPLICATION NUMBER: US/08/340,820
; ,, FILING DATE:
; ,, CLASSIFICATION: 435
; ,, PRIOR APPLICATION DATA:
; ,, APPLICATION NUMBER: US/07/835,713
; ,, FILING DATE:
; ,, ATTORNEY/AGENT INFORMATION:
; ,, NAME: CONLIN, David G.
; ,, REGISTRATION NUMBER: 27026
; ,, TELECOMMUNICATION INFORMATION:
; ,, TELEPHONE: (617)523-3400
; ,, TELEFAX: (617)523-6440
; ,, TELEX: 200291 STRE UR
; ,, INFORMATION FOR SEQ ID NO: 24:
; ,, SEQUENCE CHARACTERISTICS:
; ,, LENGTH: 1493 base pairs
; ,, TYPE: nucleic acid
; ,, STRANDEDNESS: double
; ,, TOPOLOGY: linear
; ,, MOLECULE TYPE: cDNA to mRNA
; ,, HYPOTHETICAL: NO
; ,, ANTI-SENSE: NO
; ,, ORIGINAL SOURCE:
; ,, ORGANISM: Homo sapiens
; ,, HAPLOTYPE: 2n
; ,, TISSUE TYPE: skin
; ,, CELL TYPE: fibroblast
; ,, IMMEDIATE SOURCE:
; ,, LIBRARY: Human foreskin cDNA library
; ,, CLONE: pGAF1
US-08-340-820-24

Query Match   2.6%;   Score 87.8;   DB 1;   Length 1493;
Best Local Similarity   75.9%;   Pred. No. 4.1e-09;
Matches   107;   Conservative   1;   Mismatches   33;   Indels   0;   Gaps   0;

Qy   3264 GTAAGTGAGCAATCAGAGTATAATGTTATGTGACAAAATTAAAGGCTTTCTTATATGT 3323
            |||| || | | ||| ||||| ||||||||||||| | ||| |||||||||
Db   1353 GTAAAAAATAAAAAAAAAAAAAAAAATAAAAAAAAGTTAAATTTATTTATAGAAAT 1412

Qy   3324 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3383
            |||||||||||||||||||||||||||||||||:|||||||||||||
Db   1413 TCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1472

Qy   3384 AAAAAAAAAAAAAATAAAAA 3404
            |||||||||||||| |||||
Db   1473 AAAAAAAAAAAAAAAAAAAA 1493

RESULT 12
US-08-593-535-24
; Sequence 24, Application US/08593535
; Patent No. 5622928
; ,, GENERAL INFORMATION:
; ,, APPLICANT: NARUO, Ken-ichi
; ,, APPLICANT: SEKO, Chisako
; ,, APPLICANT: KUROKAWA, Tsutomu
; ,, APPLICANT: KONDO, Tatsuya
; ,, TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
; ,, TITLE OF INVENTION: PRODUCTION
; ,, NUMBER OF SEQUENCES: 27
; ,, CORRESPONDENCE ADDRESS:
; ,, ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ,, ADDRESSEE: CUSHMAN
; ,, STREET: 130 Water Street
; ,, CITY: Boston
; ,, STATE: Massachusetts
; ,, COUNTRY: US
; ,, ZIP: 02109
; ,, COMPUTER READABLE FORM:
; ,, MEDIUM TYPE: Floppy disk
; ,, COMPUTER: IBM PC compatible
; ,, OPERATING SYSTEM: PC-DOS/MS-DOS
; ,, SOFTWARE: PatentIn Release #1.0, Version #1.25
; ,, CURRENT APPLICATION DATA:
; ,, APPLICATION NUMBER: US/08/593,535
; ,, FILING DATE: 24-JAN-1996
; ,, CLASSIFICATION: 435
; ,, PRIOR APPLICATION DATA:
; ,, APPLICATION NUMBER: US 07/835,713
; ,, FILING DATE: 12-FEB-1992
; ,, ATTORNEY/AGENT INFORMATION:
; ,, NAME: CONLIN, David G.
; ,, REGISTRATION NUMBER: 27026
; ,, TELECOMMUNICATION INFORMATION:
; ,, TELEPHONE: (617)523-3400
; ,, TELEFAX: (617)523-6440
; ,, TELEX: 200291 STRE UR
; ,, INFORMATION FOR SEQ ID NO: 24:
; ,, SEQUENCE CHARACTERISTICS:
; ,, LENGTH: 1493 base pairs
; ,, TYPE: nucleic acid
; ,, STRANDEDNESS: double
; ,, TOPOLOGY: linear
; ,, MOLECULE TYPE: cDNA to mRNA
; ,, HYPOTHETICAL: NO
; ,, ANTI-SENSE: NO
; ,, ORIGINAL SOURCE:
; ,, ORGANISM: Homo sapiens
; ,, HAPLOTYPE: 2n
; ,, TISSUE TYPE: skin
; ,, CELL TYPE: fibroblast
; ,, IMMEDIATE SOURCE:
; ,, LIBRARY: Human foreskin cDNA library
; ,, CLONE: pGAF1
US-08-593-535-24

Query Match   2.6%;   Score 87.8;   DB 1;   Length 1493;
Best Local Similarity   75.9%;   Pred. No. 4.1e-09;
Matches   107;   Conservative   1;   Mismatches   33;   Indels   0;   Gaps   0;

Qy   3264 GTAAGTGAGCAATCAGAGTATAATGTTATGTGACAAAATTAAAGGCTTTCTTATATGT 3323
            |||| || | | ||| ||||| ||||||||||||| | ||| |||||||||
Db   1353 GTAAAAAATAAAAAAAAAAAAAAAAATAAAAAAAAGTTAAATTTATTTATAGAAAT 1412

Qy   3324 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3383
            |||||||||||||||||||||||||||||||||:|||||||||||||
Db   1413 TCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1472

Qy   3324 TTAAAAAAAAAAAAAATAAAAA 3404
            |||||||||||||| |||||
Db   1413 AAAAAAAAAAAAAAAAAAAA 1493

```
Db   1473  AAAAAAAAAAAAAAAAAAAAA 1493
```

RESULT 13
US-08-473-981A-5
; Sequence 5, Application US/08473981A
; Patent No. 5629162
; GENERAL INFORMATION:
;   APPLICANT: deFougerolles, Antonin R
;   APPLICANT: Springer, Timothy A
;   TITLE OF INVENTION: METHODS OF IDENTIFYING AGENTS WHICH MODULATE
;   TITLE OF INVENTION: ICAM-3 BINDING TO LFA-1 (AS AMENDED)
;   NUMBER OF SEQUENCES: 14
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;   STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
;   CITY: WASHINGTON
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/473,981A
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: MILLONIG, ROBERT C
;   REGISTRATION NUMBER: 34,395
;   REFERENCE/DOCKET NUMBER: 1011.0560004
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 371-2600
;   TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1817 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: both
;   TOPOLOGY: both
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 9..1649
US-08-473-981A-5

  Query Match   2.5%; Score 86.6; DB 1; Length 1817;
  Best Local Similarity   85.6%; Pred. No. 7.6e-09;
  Matches 95; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

Qy   3299 CAAAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3358
          || |||| | |||| ||| | ||||||||||||||||||||||||||||||||||||||
Db   1707 CACCATAAAGCCTTCAAACTCCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1766

Qy   3359 AAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAA 3409
          |||||||| :||||||||||||||||||||||||| |||||||||||
Db   1767 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1817
```

RESULT 14
PCT-US95-06406A-21
; Sequence 21, Application PC/TUS9506406A
; GENERAL INFORMATION:
;   APPLICANT: Janet D. Robishaw, Charles Kunsch
;   TITLE OF INVENTION: cDNA Clones Encoding Human G Protein
;   TITLE OF INVENTION: Subunits
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:

;   STREET:
;   CITY:
;   STATE:
;   COUNTRY:
;   ZIP:
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;   COMPUTER: IBM 486
;   OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
;   SOFTWARE: WORDPERFECT 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/06406A
;   FILING DATE: Herewith
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME:
;   REGISTRATION NUMBER:
;   REFERENCE/DOCKET NUMBER:
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:
;   TELEFAX:
; INFORMATION FOR SEQ ID NO: 21:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 903
;   TYPE: Nucleic Acid
;   STRANDEDNESS: Single
;   TOPOLOGY: Linear
;   ANTI-SENSE: No
PCT-US95-06406A-21

  Query Match   2.5%; Score 85; DB 5; Length 903;
  Best Local Similarity 84.7%; Pred. No. 1.2e-08;
  Matches 94; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy   3300 AAATTAAAGGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3359
          || | |||| ||| | |||||||||||||||||||||||||||||||||
Db   788 AAGAATAAAGTCATCCAGAGCCTCAAAAAAAAAAAAAAAAAAATAAAAAAAAAA 847

Qy   3360 AAAAAAAARAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAA 3410
          |||||||| :||||||||||||||||||||||||| |||||||||||
Db   848 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 898
```

RESULT 15
US-08-785-310A-2
; Sequence 2, Application US/08785310A
; Patent No.5840532
; GENERAL INFORMATION:
;   APPLICANT: McKnight, Steven L.
;   APPLICANT: Russell, David W.
;   TITLE OF INVENTION: Neuronal PAS Domain Protein
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
;     STREET: 268 BUSH STREET, SUITE 3200
;     CITY: SAN FRANCISCO
;     STATE: CALIFORNIA
;     COUNTRY: USA
;     ZIP: 94104
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/785,310A
;     FILING DATE: 21-JAN-1997
;     CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:

;   NAME: OSMAN, RICHARD A
;   REGISTRATION NUMBER: 36,627
;   REFERENCE/DOCKET NUMBER: UTSD:1226
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 343-4341
;     TELEFAX:   (415) 343-4342
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2082 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: CDNA
US-08-785-310A-2

Query Match       2.5%;  Score 84;  DB 4;  Length 2082;
Best Local Similarity   88.2%;  Pred. No. 2.6e-08;
Matches  90;  Conservative   1;  Mismatches  11;   Indels    0;   Gaps    0;

Qy  3309 GGCTTTCTTATATGTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3368
         | || || | | |   ||||||||||||||||||||||||||||||||||||||
Db  1953 GCCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2012

Qy  3369 RAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA 3410
         :|||||||||||||||||||||||||||| ||||||||||||||
Db  2013 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2054

Search completed: September 25, 1999, 07:29:01
Job time: 3378 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on:    September 25, 1999, 09:49:56 ; Search time 1811.29 Seconds
(without alignments)
3713.563 Million cell updates/sec

Title:    US-09-030-606-110
Perfect score:    3410
Sequence:    1 GGGAACCAGCCTGCACGCGC.........AAAAAAATAAAAAAAAAAAA 3410

Scoring table:    IDENTITY_NUC
Searched:    2546578 seqs, 986266752 residues
Database :    EST:*

| | | |
|---|---|---|
| 1: em_est1:* | 2: em_est2:* | 3: em_est3:* |
| 4: em_est4:* | 5: em_est5:* | 6: em_est6:* |
| 7: em_est7:* | 8: em_est8:* | 9: em_est9:* |
| 10: em_est10:* | 11: em_est11:* | 12: em_est12:* |
| 13: em_est13:* | 14: em_est14:* | 15: em_est15:* |
| 16: em_est16:* | 17: em_est17:* | 18: em_est18:* |
| 19: em_est19:* | 20: gb_est1:* | 21: gb_est2:* |
| 22: gb_est3:* | 23: gb_est4:* | 24: gb_est5:* |
| 25: gb_est6:* | 26: gb_est7:* | 27: gb_est8:* |
| 28: gb_est9:* | 29: gb_est10:* | 30: gb_est11:* |
| 31: gb_est12:* | 32: gb_est13:* | 33: gb_est14:* |
| 34: gb_est15:* | 35: gb_est16:* | 36: gb_est17:* |
| 37: gb_est18:* | 38: gb_est19:* | 39: gb_est20:* |
| 40: gb_est21:* | 41: gb_est22:* | 42: gb_est23:* |
| 43: gb_est24:* | 44: gb_est25:* | 45: gb_est26:* |
| 46: gb_est27:* | 47: gb_est28:* | 48: gb_est29:* |
| 49: gb_est30:* | 50: gb_est31:* | 51: gb_est32:* |
| 52: em_est20:* | 53: em_est21:* | 54: em_est22:* |
| 55: em_est23:* | 56: em_est24:* | 57: em_est25:* |
| 58: em_est26:* | | |

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| c | 1 | 456.8 | 13.4 | 491 | 36 | AA631143 | AA631143 nq75g08.s |
| c | 2 | 449.2 | 13.2 | 469 | 50 | AI703348 | AI703348 wd93b09.x |
| c | 3 | 366 | 10.7 | 388 | 35 | AA579486 | AA579486 nf33g07.s |
| c | 4 | 333.8 | 9.8 | 348 | 40 | AA984323 | AA984323 am84a12.s |
| c | 5 | 324.2 | 9.5 | 375 | 30 | AA225106 | AA225106 nc21d11.r |
| c | 6 | 314 | 9.2 | 315 | 36 | AA640153 | AA640153 np28b03.s |
| c | 7 | 311 | 9.1 | 313 | 48 | AI587483 | AI587483 tr51c10.x |
| c | 8 | 306 | 9.0 | 306 | 36 | AA631024 | AA631024 nq76g11.s |
| c | 9 | 300.2 | 8.8 | 674 | 47 | AI525162 | AI525162 promrna-9 |
| c | 10 | 295.6 | 8.7 | 346 | 25 | N95796 | N95796 zb66h02.s1 |
| c | 11 | 282.6 | 8.3 | 301 | 50 | AI696721 | AI696721 wc56d05.x |
| c | 12 | 281.8 | 8.3 | 299 | 47 | AI468280 | AI468280 tg57a01.x |
| c | 13 | 280.6 | 8.2 | 342 | 28 | AA112573 | AA112573 zm28c12.r |
| c | 14 | 262.4 | 7.7 | 264 | 34 | AA492342 | AA492342 ng81d12.s |
| c | 15 | 255.6 | 7.5 | 288 | 35 | AA579735 | AA579735 nf39g07.s |
| c | 16 | 253.4 | 7.4 | 287 | 35 | AA570251 | AA570251 nf39d12.s |
| c | 17 | 242 | 7.1 | 242 | 47 | AI472247 | AI472247 t177a04.x |
| c | 18 | 228 | 6.7 | 228 | 28 | AA112574 | AA112574 zm28c12.s |
| c | 19 | 221.4 | 6.5 | 239 | 26 | W24907 | W24907 zb66h02.r1 |
| c | 20 | 214.8 | 6.3 | 253 | 35 | AA552457 | AA552457 nj90f02.s |
| c | 21 | 214 | 6.3 | 214 | 36 | AA652651 | AA652651 ns64c11.s |
| c | 22 | 165.4 | 4.9 | 363 | 42 | AU023209 | AU023209 AU023209 |
| c | 23 | 162.8 | 4.8 | 242 | 36 | AA647708 | AA647708 vq77h04.s |
| c | 24 | 160 | 4.7 | 173 | 36 | AA652452 | AA652452 ns64e11.s |
| c | 25 | 142.4 | 4.2 | 163 | 35 | AA579320 | AA579320 nf36e12.s |
| c | 26 | 123.6 | 3.6 | 686 | 48 | AI598307 | AI598307 EST250010 |
| c | 27 | 120 | 3.5 | 378 | 42 | AI137795 | AI137795 UI-R-C0-h |
| c | 28 | 118 | 3.5 | 320 | 46 | AA998873 | AA998873 UI-R-C0-h |
| c | 29 | 117.8 | 3.5 | 282 | 46 | AA957294 | AA957294 UI-R-E1-f |
| c | 30 | 102.6 | 3.0 | 550 | 29 | AA137485 | AA137485 mq98h06.r |
| c | 31 | 101.2 | 3.0 | 559 | 39 | C88248 | C88248 C88248 Mous |
| c | 32 | 101.2 | 3.0 | 573 | 39 | C88345 | C88345 C88345 Mous |
| c | 33 | 101.2 | 3.0 | 556 | 42 | AU022931 | AU022931 AU022931 |
| c | 34 | 101.2 | 3.0 | 589 | 42 | AU023994 | AU023994 AU023994 |
| c | 35 | 101.2 | 3.0 | 411 | 44 | AU043413 | AU043413 AU043413 |
| c | 36 | 101.2 | 3.0 | 438 | 44 | AU043506 | AU043506 AU043506 |
| c | 37 | 101.2 | 3.0 | 459 | 44 | AU043764 | AU043764 AU043764 |
| c | 38 | 101.2 | 3.0 | 462 | 44 | AU043793 | AU043793 AU043793 |
| c | 39 | 101.2 | 3.0 | 566 | 44 | AU044322 | AU044322 AU044322 |
| c | 40 | 101.2 | 3.0 | 360 | 44 | AU044508 | AU044508 AU044508 |
| c | 41 | 101.2 | 3.0 | 297 | 44 | AU045544 | AU045544 AU045544 |
| c | 42 | 100 | 2.9 | 349 | 39 | C86407 | C86407 C86407 Mous |
| c | 43 | 99.2 | 2.9 | 457 | 42 | AU024466 | AU024466 AU024466 |
| c | 44 | 98.8 | 2.9 | 185 | 35 | AA589036 | AA589036 v163d01.r |
| c | 45 | 98.6 | 2.9 | 285 | 47 | AI536638 | AI536638 to14e05.x |

ALIGNMENTS

RESULT 1
AA631143/c
LOCUS    AA631143    491 bp    mRNA    EST    31-OCT-1997
DEFINITION    nq75g08.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1158206 3',
mRNA sequence.
ACCESSION    AA631143
NID    g2553754
VERSION    AA631143.1    GI:2553754

```
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 491)
AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       On Sep 12, 1996 this sequence version replaced gi:1400896.

              Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
              Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html

              Insert Length: 919   Std Error: 0.00
              Seq primer: -40m13 fwd. ET from Amersham
              High quality sequence stop: 472.
FEATURES          Location/Qualifiers
     source       1..491
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:1158206"
                  /clone_lib="NCI_CGAP_Pr22"
                  /sex="male"
                  /tissue_type="normal prostate"
                  /lab_host="DH10B"
                  /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                  with a modified polylinker; 1st strand cDNA was prepared
                  from normal prostate bulk tissue, and was then primed with
                  a Not I - oligo(dT) primer. Double-stranded cDNA was
                  ligated to Eco RI adaptors (Pharmacia), digested with Not
                  I and cloned into the Not I and Eco RI sites of the
                  modified pT7T3 vector. Library is normalized, and was
                  constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      135 a    104 c    136 g    116 t
ORIGIN

Query Match         13.4%;   Score 456.8; DB 36;  Length 491;
Best Local Similarity 98.0%;  Pred. No. 8.8e-70;
Matches 484;   Conservative   0;  Mismatches   0;  Indels   3;  Gaps   2;

Qy  2837 CAAGGTTAGGGTGTTGAAGGAAGGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCCTA 2896
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   491 CAAGGTTAGGGTGTTGAAGGAAGGTAGAGGGTGGGGCTTCAGGTCTCAACGGCTTCCCTA 432

Qy  2897 ACCACCCCCTCTTCCTCTTGGCCCAGCCTGGTTCCCCGCCCACTTCCACTCCCCTACTCTCT 2956
         ||||||||||||||| ||||  |||||||||||||||||||||||||||||||||||||
Db   431 ACCACCCCCTCTTTCTG--CCCAGCCGTGTCCCCTACTTCCACTCCCCTCTACTCTCT 374

Qy  2957 CTAGGACTGGGCTGATGAAGGCACTGCCCAAAATTTCCCCTACCCCACCCCAACTTCCCCTAC 3016
         ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   373 CTAGGACTGGGCTGATGAA-GCACTGCCCAAAATTTCCCCTACCCCACCCCAACTTCCCCTAC 315

Qy  3017 CCCAACTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCAC 3076
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 CCCAACTTCCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCAC 255

Qy  3077 AAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGCTTGGGG 3136
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254 AAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGCTTGGGG 195

Qy  3137 AATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTGTTTATCTCTCA 3196
```

```
                                                                  135 t
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 AATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTAATCTCTCA 135

Qy  3197 GGGGGGTTTAAGTGCCGTTGCCGTTGCAATAATGTCGTCTTATTATTTAGCGGGTGAATATT 3256
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134 GGGGGGGTTTAAGTGCCGTTGCCGTTGCAATAATGTCGTCTTATTTATTTAGCGGGTGAATATT 75

Qy  3257 TTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTCT 3316
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    74 TTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTCT 15

Qy  3317 TATATGTTTAAAAA 3330
         ||||||||||||||
Db    14 TATATGTTTAAAAA 1


RESULT  2
AI703348/c
LOCUS       AI703348          469 bp    mRNA            EST           03-JUN-1999
DEFINITION  wd93b09.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2339129 3',
            mRNA sequence.
ACCESSION   AI703348
NID         g4991248
VERSION     AI703348.1  GI:4991248
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 469)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3188436.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 447.
FEATURES          Location/Qualifiers
     source       1..469
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:2339129"
                  /clone_lib="NCI_CGAP_Lu24"
                  /tissue_type="carcinoid"
                  /lab_host="DH10B"
                  /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                  modified polylinker; Plasmid DNA from the normalized
                  library NCI_CGAP_Lu5 was prepared, and ss circles were
                  made in vitro. Following HAP purification, this DNA was
                  used as tracer in a subtractive hybridization reaction.
                  The driver was PCR-amplified cDNAs from a pool of 5,000
                  clones made from the same library (cloneIDs
                  1414920-1417991 and 1520904-1522439). Subtraction by Bento
                  Soares and M. Fatima Bonaldo. "           4 others
BASE COUNT      126 a     97 c    135 g    107 t

ORIGIN

Query Match         13.2%;   Score 449.2; DB 50;  Length 469;
Best Local Similarity 98.5%;  Pred. No. 1.8e-68;
```

```
Matches 462;  Conservative   0;  Mismatches   6;  Indels   1;  Gaps   1;

Qy  2857  AAGTAGAGGTGGGGCTCAGGTCTCAACGGCTTCCCTAACCACCCCTCTTCTTGGC  2916
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   469  AAGTAGAGGTGGGGCTTCAGGTCTCAACGGCTTCCCTAACCACCCCTTCTTGGC   410

Qy  2917  CCAGCCTGGTTCCCCCACTTCCACTCCCTCTACTCTCTAGGACTGGGCTGATGAAG  2976
          ||||||||||||||||||||||||||  |||||| || |||||| ||  |||||
Db   409  CCAGCCTGGTTCCCCCACTTCCACTCCCTCTACTCTCATCAGGGACTGGNCTNATGAAG   350

Qy  2977  GCACTGCCCAAAATT-CCCTACCCCCAACTTTCCCCACCAG  3035
          |||||| ||||||||  |||||||||||||||||||||||||
Db   349  GCACTGCNCAAAATTNCCCTACCCCCAACTTTCCCCACCAG   290

Qy  3036  CTCCACAACCCTGTTGGAGCTACTGCAGGACCACAAGGTGCGGTTCCCAAG  3095
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   289  CTCCACAACCCTGTTGGAGCTACTGCAGGACCACAAGTGCGGTTCCCAAG   230

Qy  3096  CTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAGAAACTCA  3155
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   229  CTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAGAAACTCA   170

Qy  3156  GGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGT  3215
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   169  GGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGT   110

Qy  3216  TTGCAATAATGTCGTCTTATTTATTTAGGCGGGTGAATATTTTATACTGTAAGTGAGCAA  3275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   109  TTGCAATAATGTCGTCTTATTTATTTAGGCGGGTGAATATTTTATACTGTAAGTGAGCAA    50

Qy  3276  TCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTTATATGTT  3324
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    49  TCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTTATATGTT     1
```

```
RESULT 3
AA579486/c
LOCUS       AA579486        388 bp    mRNA      EST       03-SEP-1997
DEFINITION  nf33g07.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:915612
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AA579486
NID         g2357670
VERSION     AA579486.1  GI:2357670
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 388)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On May 8, 1995 this sequence version replaced gi:800271.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 343.
FEATURES    Location/Qualifiers
source      1..388
            /organism="Homo sapiens"
```

```
            /db_xref="taxon:9606"
            /clone="IMAGE:915612"
            /clone_lib="NCI_CGAP_Pr1"
            /sex="Male"
            /dev_stage="45 years old"
            /lab_host="DH10B"
            /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
            strand cDNA was primed with oligo(dT)17 on 50 ng of
            DNase-treated, total cellular RNA obtained from
            5,000-10,000 microdissected, histologically normal
            prostate epithelial cells. Double-stranded cDNA was
            ligated to EcoRI adaptors, 5 cycles of PCR applied to the
            cDNA with an adaptor-specific primer, and the resulting
            PCR product subcloned into pAMP10 by the UDG-cloning
            method (Life Technologies). Average insert size is 600
            bp. NOTE: Not directionally cloned. This library was
            constructed by David Krizman."
```

```
BASE COUNT     99 a      77 c      99 g     113 t
ORIGIN
```

```
Query Match                10.7%;    Score 366;  DB 35;  Length 388;
Best Local Similarity      98.7%;    Pred. No. 4e-54;
Matches 369;  Conservative   0;  Mismatches   5;  Indels   0;  Gaps   0;

Qy  2975  AGGCACTGCCCAAAATTCCCCTACCCCCAACTTTCCCCACCA  3034
          | ||||||||||||||||||||||||||||||||||||||||
Db   381  AAGCACTGCCCAAAATTCCCCTACCCCCAACTTTCCCCACCA   322

Qy  3035  GCTCCACAACCCTGTTGGAGCTACTGCAGGACCACAAGTGCGGTTCCCAAG  3094
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321  GCTCCACAACCCTGTTGGAGCTACTGCAGGACCACAAGTGCGGTTCCCAAG   262

Qy  3095  CCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAGAAACTC  3154
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261  CCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAGAAACTC   202

Qy  3155  AGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCG  3214
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201  AGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCG   142

Qy  3215  TTTGCAATAATGTCGTCTTATTTATTTAGGCGGGTGAATATTTTATACTGTAAGTGAGCA  3274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141  TTTGCAATAATGTCGTCTTATTTATTTAGGCGGGTGAATATTTTATACTGTAAGTGAGCA    82

Qy  3275  ATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTTATATGTTTAAAAAAAAA  3334
          |||||||||||||||||||||||||||||||||||||||||           |||||||||
Db    81  ATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTTAACCGTTTAAAAAAAAA    22

Qy  3335  AAAAAAAAAAAAA  3348
          |||||||||||||
Db    21  AAAAAAAAAAAAAA     8
```

```
RESULT 4
AA984323/c
LOCUS       AA984323        348 bp    mRNA      EST       27-MAY-1998
DEFINITION  am84a12.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone
            IMAGE:1629790 3', mRNA sequence.
ACCESSION   AA984323
NID         g3162848
VERSION     AA984323.1  GI:3162848
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 348)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
```

JOURNAL   Unpublished (1997)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1402298.

             Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Possible reversed clone: polyT not found
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 348.

FEATURES   Location/Qualifiers
        source   1..348
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1629790"
                     /clone_lib="Stratagene schizo brain S11"
                     /sex="male"
                     /tissue_type="schizophrenic brain s-11 frontal lobe"
                     /dev_stage="34 years old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Library
                     constructed from S-11 frontal lobe, male, 34 years old,
                     50% caucasian, 50% Aleutian. Schizophrenic suicide.
                     Random primed into EcoRI site of ZAP II Vector. Mass
                     excised. Custom library. Avg insert length 1.4kb.
                     Material obtained by Johnston N., Torrey, E.F., Yolken R.,
                     and the Stanley Neuropathology Consortium - Analysis of
                     RNAs from the Brains of Individuals with Psychiatric
                     Diseases (Unpublished) Stanley Neurovirology Laboratory,
                     Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT   78 a   99 c   118 g   53 t
ORIGIN

Query Match   9.8%;   Score 333.8;   DB 40;   Length 348;
Best Local Similarity   99.1%;   Pred. No. 1.4e-48;
Matches 346;   Conservative   0;   Mismatches 2;   Indels   1;   Gaps   1;

Qy   1259 CTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTGGTCATGGACCGGCTG 1318
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   348 CTGGGGCTGTTCCTGCAGTGCGCCATCTCCCTGGTCTTCTCTGGTCATGGACCGGCTG 289

Qy   1319 GTGCAGGGATTCGGCACTCGAGCAGTCTATTGGCCAGTGTGCGCAGTTTCCTGTGGCT 1378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   288 GTGCAGGGATTCGGCACTCGAGCAGTCTATTGGCCAGTGTGCGCAGTTTCCTGTGGCT 229

Qy   1379 GCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGTGACAGCTTCAGCCGCCCTCACC 1438
          |||||||||||||||||||||||||||||||||||||||||||| |||||
Db   228 GCCGGTGCCACATGCCTGTCCCACAGTGTGGCCGTGTGACAGCTTCAGCCGACCTCACC 169

Qy   1439 GGGTTCACCTTCTCAGCCTTCAGCCTGCAGATCCTGCCCTCGCCTCCCTCTACCACCGG 1498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   168 GGGTTCACCTTCTCAGCCTTCAGCCTGCAGATCCTGCCCTCGCCTCCCTCTACCACCGG 109

Qy   1499 GAGAAGCAGGTGTGTTCCTGCCCAAATACCGAGGGACACTGGAGGTGCTAGCAGTGAGGAC 1558
          |||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   108 GAGAAGCAGGTGTGTTCCTGCCCAAATACCGAGGGGACACTGGAGGTGCTGGCAGTGAGGAC 49

Qy   1559 AGCCTGATGACCAGCTTCCTGCCCAGGCCTAAGCCTGGAGCTCCCTTCC 1607
          ||||||||||||||||||||||||| |||||||||||||||||||||||
Db   48 AGCCTGATGACCAGCTTCCTG-CAGGCCCTAAGCCTGGAGCTCCCTTCC 1

RESULT   5
AA225106   AA225106   375 bp   mRNA   EST   15-AUG-1997
LOCUS   nc21d11.r1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:1008789, mRNA
DEFINITION   sequence.
ACCESSION   AA225106

NID   gi1846415
VERSION   AA225106.1   GI:1846415
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 375)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Apr 14, 1993 this sequence version replaced gi:692639.

             Contact: Robert Strausberg, Ph.D.
             Tel: (301) 496-1550
             Email: Robert_Strausberg@nih.gov
             Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
             M.D., Michael Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html

             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 305.

FEATURES   Location/Qualifiers
        source   1..375
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1008789"
                     /clone_lib="NCI_CGAP_Pr1"
                     /sex="Male"
                     /dev_stage="45 years old"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was primed with oligo(dT)17 on 50 ng of
                     DNAse-treated, total cellular RNA obtained from
                     5,000-10,000 microdissected, histologically normal
                     prostate epithelial cells. Double-stranded cDNA was
                     ligated to EcoRI adaptors, 5 cycles of PCR applied to the
                     cDNA with an adaptor-specific primer, and the resulting
                     PCR product subcloned into pAMP10 by the UDG-cloning
                     method (Life Technologies). Average insert size is 600
                     bp. NOTE: Not directionally cloned. This library was
                     constructed by David Krizman."
BASE COUNT   98 a   94 c   76 g   101 t   6 others
ORIGIN

Query Match   9.5%;   Score 324.2;   DB 30;   Length 375;
Best Local Similarity   97.0%;   Pred. No. 6.7e-47;
Matches 359;   Conservative   0;   Mismatches 8;   Indels   3;   Gaps   3;

Qy   2966 GGCTGATGAAGGCACTGCCCAAAATTCCCCTACCCCAACTTCCCTACCCCCAACTT 3025
          | |||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   6 GNCTGATGAAGGCACTGCCCAAAATTNCCCTANCCCCAACTTCCCTACCCCCAACTT 65

Qy   3026 TCCCCACCAGCTCCACAACCCTGTTGGAGCTACTGGAGCCAGGAGCACAAAGTGCGG 3085
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 TCCCCACCAGCTCCACAACCCTGTTGGAGCTACTGGAGCCAGGAGCACAAAGTGGCT 125

Qy   3086 TTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACA 3145
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 NTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGAATCTCACA 185

Qy   3146 CAGAAACTCAGGAGCAGCACCCCCTGCC-TGAGCTAAGGAGGGTTCTATCTCCAGGGGGGT 3204
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   186 CAGAAACTCAGGAGGAGCACCCCCGTGGCTGAGCTTATCTCTCAGGGGGGT 245

Qy   3205 TTAAGTGCCGTTGCAATAATGTCGTCTTATTTATTTAGC-GGGGTGAATATTTTATACT 3263

```
Db   315  CCCTACCCCCAACTTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAG  256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  3071  AAGCACAAAGTGCGGTTCCCAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC  3130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255  AAGCACAAAGTGCGGTTCCCAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGC  196
          
Qy  3131  TTGGGGAATCTCACACAGAAACTCAGGAGCACCACCCCCTGCCTAGGGAGGTCTTAT  3190
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195  TTGGGGAATCTCACACAGAAACTCAGGAGCACCACCCCCTGCCTAGGGAGGTCTTAT  136
          
Qy  3191  CCCTCAGGGGGGTTAAGTGCCGTTGCAATAATGTCGTCTTATTATTTAGCGGGGTG  3250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135  CCCTCAGGGGGGTTAAGTGCCGTTGCAATAATGTCGTCTTATTATTTAGCGGGGTG  76
          
Qy  3251  AATATTTTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGG  3310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75  AATATTTTATACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGG  16
          
Qy  3311  CTTTCTTATATGTT  3324
          ||||||||||||||
Db    15  CTTTCTTATATGTT  2
```

```
RESULT    7
AI587483/c
LOCUS       AI587483            313 bp    mRNA            EST       14-MAY-1999
DEFINITION  tr51c10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2221842 3',
            mRNA sequence.
ACCESSION   AI587483
NID         g4573924
VERSION     AI587483.1  GI:4573924
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189383.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Life Technologies catalog #: 11548-013
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 1960   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 310
            POLYA=NO.
FEATURES             Location/Qualifiers
     source          1..313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2221842"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT    87 a     69 c     76 g     81 t
ORIGIN
Query Match       9.1%;  Score 311;  DB 48;  Length 313;
```

```
Db   246  TTAAGTGCCGTTGCAATAATGTCGTCTTATTATTTAGCNGGGTGAATATTTTATACT  305
          ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Qy  3264  GTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATT-AAAGGCTTTCTTATATG  3322
          |||||||||||||||||||||||||||||||||||||||||  |||||||||||||||
Db   306  GTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAAGGCTTTCTTATATG  365
          
Qy  3323  TTTAAAAAAA  3332
          ||||| ||||
Db   366  TTTAAANAAA  375
```

```
RESULT    6
AA640153/c
LOCUS       AA640153            315 bp    mRNA            EST       23-OCT-1997
DEFINITION  np28b03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1117613 3',
            mRNA sequence.
ACCESSION   AA640153
NID         g2563932
VERSION     AA640153.1  GI:2563932
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 315)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:430583.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1..315
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="21"
                     /clone="IMAGE:1117613"
                     /clone_lib="NCI_CGAP_Pr22"
                     /sex="male"
                     /tissue_type="normal prostate"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; 1st strand cDNA was prepared
                     from normal prostate bulk tissue, and was then primed with
                     a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT    89 a     69 c     77 g     80 t
ORIGIN
Query Match       9.2%;  Score 314;  DB 36;  Length 315;
Best Local Similarity 100.0%;  Pred. No. 3.7e-45;
Matches 314;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy  3011  CCCTACCCCCAACTTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAG  3070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      Best Local Similarity 100.0%; Pred. No. 1.2e-44;
      Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  3012 CCTACCCCCAACTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGA 3071
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 CCTACCCCCAACTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGA 254

Qy  3072 AGCACAAAGTGCGGTTCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCT 3131
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 AGCACAAAGTGCGGTTCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCT 194

Qy  3132 TGGGGAATCCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATC 3191
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 TGGGGAATCCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATC 134

Qy  3192 TCTCAGGGGGGTTAAGTGCCGTTGCAATAATGTCGTCTTATTTATTTAGCGGGGGTGA 3251
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 TCTCAGGGGGGGTTAAGTGCCGTTGCAATAATGTCGTCTTATTTATTTAGCGGGGGTGA 74

Qy  3252 ATATTTTATACTGTAAGTGAGCAATCAGAGTATATATGTTTATGGTGACAAAATTAAAGGC 3311
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    73 ATATTTTATACTGTAAGTGAGCAATCAGAGTATATATGTTTATGGTGACAAAATTAAAGGC 14

Qy  3312 TTTCTTATATG 3322
         |||||||||||
Db    13 TTTCTTATATG 3

RESULT      8
AA631024/c
LOCUS       AA631024   306 bp   mRNA        EST      31-OCT-1997
DEFINITION  nq76g11.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1158308 3',
            mRNA sequence.
ACCESSION   AA631024
NID         g2553635
VERSION     AA631024.1  GI:2553635
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 306)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Sep 12, 1996 this sequence version replaced gi:1397799.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 375   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES    Location/Qualifiers
     source 1..306
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1158308"
            /clone_lib="NCI_CGAP_Pr22"
            /sex="male"
            /tissue_type="normal prostate"
            /lab_host="DH10B"
            /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
            with a modified polylinker; 1st strand cDNA was prepared
```

```
            from normal prostate bulk tissue, and was then primed with
            a Not I - oligo(dT) primer. Double-stranded cDNA was
            ligated to Eco RI adaptors (Pharmacia), digested with Not
            I and cloned into the Not I and Eco RI sites of the
            modified pT7T3 vector.  Library is normalized, and was
            constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  89 a   69 c   70 g   78 t
ORIGIN

      Query Match     9.0%;  Score 306;  DB 36;  Length 306;
      Best Local Similarity 100.0%; Pred. No. 8.9e-44;
      Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  3020 CAACTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAGCACAAA 3079
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306 CAACTTCCCACCAGCTCCACAACCCTGTTTGGAGCTACTGCAGGACCAGAAGCACAAA 247

Qy  3080 GTGCGGTTCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATCTGTCTTGGGGAAT 3139
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246 GTGCGGTTCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATCTGTCTTGGGGAAT 187

Qy  3140 CTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGG 3199
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186 CTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTCAGGG 127

Qy  3200 GGGGTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATTTATTTAGCGGGGGTGAATATTTTA 3259
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 GGGGTTAAGTGCCGTTTGCAATAATGTCGTCTTATTTATTTATTTAGCGGGGGTGAATATTTTA 67

Qy  3260 TACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTCTTAT 3319
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66 TACTGTAAGTGAGCAATCAGAGTATAAATGTTTATGGTGACAAAATTAAAGGCTTTCTTAT 7

Qy  3320 ATGTTT 3325
         ||||||
Db     6 ATGTTT 1

RESULT      9
AI525162
LOCUS       AI525162   674 bp   mRNA        EST      18-MAR-1999
DEFINITION  promrna-9.C08.r bvtumor Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   AI525162
NID         g4439297
VERSION     AI525162.1  GI:4439297
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 674)
AUTHORS     Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun
            Yu,J. and Hood,L.
TITLE       Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3136548.

            Contact: Guyang Matthew Huang
            Leroy Hood
            University of Washington
            Department of Molecular Biotechnology, Box 357730, University of
            Washington, Seattle, WA 98195
            Tel: 5106280100
            Fax: 5106280108
            Email: huanggm@yahoo.com.
FEATURES    Location/Qualifiers
     source 1..674
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="21"
            /clone_lib="bvtumor"
            /note="Organ: Prostate; Vector: pBluescript; Directional
```

```
                    cDNA library was constructed using Lambda zP II kit
                    (Stratagene). mRNA was extracted from a fresh prostate
                    tumor tissues (Urology Department, University of
                    Washington)."

BASE COUNT     125 a     124 c     164 g     178 t      83 others
ORIGIN

Query Match              8.8%;    Score 300.2; DB 47;  Length 674;
Best Local Similarity   77.2%;    Pred. No. 1e-42;
Matches  447; Conservative   0; Mismatches 120;  Indels  12;  Gaps  10;

Qy  1919 AGGGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATGGGGTGGAGGGCCTGC 1978
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16 AGGGACTTGGCCAAATACTCAGCGTAGAAAACTTCCAGCACATGGGGTGGAGGGCCTGC 75

Qy  1979 CTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCATGGGGCTGCCGGGCTGGCCGCCA 2038
         ||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db    76 CTCACTGGGTCCCAGCTCCCCGCTCCTGTTAGCCCCATGGGGCTNGCCGGGCTGGCCGCCA 135

Qy  2039 GTTCTGTTGCTGCCAAAGTAATGGGCTTCTCTGCGCCACCCTGCTGCTGAGGTCG 2098
         ||||||||||||||||||||||||||| |||||||| |||||||||||||||||
Db   136 GTTTCTGTTGCTGCGCCAAAGTAATGGGCTNTCTGCGCCACCNTGCTGCTGAGGTGCG 195

Qy  2099 TAGCTGCACAGCTGCGGGGCTGGGGCGTCCCTCCCTCTCCCCAGTCTCTAGGGCTGCC 2158
         || -|||||||||||||| ||||| | ||||||||||||| |||||| |||||||||
Db   196 TA--TGCACAGCTGGGNGCTGNGNCGTCCCTNTCCTCTNCCCAGTNTCTAGGGCTGNC 253

Qy  2159 TGACTGGAGGCCTTCCAAGGGGTTCAGTCT-GGACTTATACAGGGAGGCCAGAAGGG- 2217
         ||||||||| || |||| |||||||||||| ||||||||||| |||||| |||| ||
Db   254 TGACTGGAGGNCTTNCAAGTGGNTTCAGTCNNGACTTATACAGNGAGGCCAGTTGGGT 313

Qy  2217 CTCCATGCACT-GGAATGCGGGGACTCTGCAGGTTGGATT-ACCCAGGCTCAGGGTTAAC- 2274
         ||||||||||| | ||||| |||||||||||||||| |  |||||||||||||||||||
Db   314 TTCCATGCACTGGGGAATGCGGGGACTTTGCAGTTGGATTAACCCCAGGNTCAGGGGTAANA 373

Qy  2274 AGCTAGCCCTCCTAGTTGAGAC-ACACCTAGAGAGAGGGTTTT-GGGAGCTGAATAAACTC 2331
         || |||||||||||||||||| | ||||||||||||||||| ||||||||||||||||||
Db   374 AGTTAGCCCTCCTAGTNGNGACACACNTAGTGAAGGGTTTTTNGGGNGCTNAATAAACTG 433

Qy  2332 AGTTCACCCTGTTTCCCCATCTCT--AAGCCCCTTAACCTGCAGCTTCGTTTAATGTAGCTC 2389
         ||||| |||| | ||| | |||  |||||||||||||||||||  |||| | ||||
Db   434 AGTNACCTGNGTNTCCANTTTNAAAGNCNTTTAATCTGNAAGTNTNTNTTAANTTNAGGT 493

Qy  2390 TTGCCATGGGAGTTTCTAGGATGAGAAACACTCCTCCATGGGATTTGAACATATGACTTATTT 2449
         || |||||||||||| || ||| || |||| |  |||| | | |||||   ||| ||||
Db   494 NTNTNAATNGGGGTTTTTAGGTGNAANATNNNNNNGGGTTG-ANANTTNANGTATTT 552

Qy  2450 GTAGGGGAAGAGTCCTGAGGGGGCAACACACAAGAACCAG 2488
         |||| ||||| | || |  | |||| || |||
Db   553 TTAGGNNANGAGCCTTNGGNANANANNNNCCAGGCCNNAG 591
```

RESULT 10
N95796/c
LOCUS   N95796   346 bp   mRNA   EST   20-AUG-1996
DEFINITION zb66h02.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
            IMAGE:308595 3', mRNA sequence.
ACCESSION   N95796
NID   g1268140
VERSION   N95796.1 GI:1268140
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 346)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.,

and Marra,M.
  TITLE   Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
  COMMENT   On May 8, 1995 this sequence version replaced gi:801262.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 622   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 263.
FEATURES   Location/Qualifiers
     source   1..346
                     /organism="Homo sapiens"
                     /db_xref="GDB:1252008"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:308595"
                     /clone_lib="Soares_fetal_lung_NbHL19W"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGAGCGGCCGCCAATTTTTTTTTTTTTTTT-3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. This library was constructed
                     from the same fetus as the fetal heart library, Soares
                     fetal heart NbHH19W."
BASE COUNT   99 a   69 c   87 g   87 t   4 others
ORIGIN

```
Query Match              8.7%;    Score 295.6; DB 25;  Length 346;
Best Local Similarity   92.5%;    Pred. No. 5.7e-42;
Matches  320; Conservative   0; Mismatches  23;  Indels   3;  Gaps   1;

Qy  2983 CCAAAAATTTCCCCACCCCCAACTTCCCTACCCCACCTTCCCCACCAGC---TCC 3039
         || ||||||||||||| |||||| | ||  | ||| ||||||||||||
Db   346 CCAAAATTTTCCCNTACCCCNAAACTTNCCCTACCCCANTTNTCCCCACCAGGTTCCA 287

Qy  3040 ACAACCCTGTTTGGAGCTACTGCAGGAGCACAGAAGCACAAAGTGCGGTTCCCAAGCCTTT 3099
         | |||||||||||||||||| ||||||||| || || ||||||||||| ||||||||||
Db   286 CAACCCTGTTTGGAGCTACTGCAGGAGCAACACCAAAGTGCGGTTCCCAAGCCTTT 227

Qy  3100 GTCCATCTCAGCCCCCAGAGTATATCTGTCTTGGGGAATCTCACACAGAAACTCAGGAG 3159
         ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   226 GTCCATCTCAGCCCCCAGAGTATATCTGTCTTGGGGAATCTCACACAGAAACTCAGGAG 167

Qy  3160 CACCCCTGCTGAGCTAAGGGAGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGC 3219
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   166 CACCCCTGCTGAGCTAAGGGAGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGC 107

Qy  3220 AATAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAAGTGAGCAATCAG 3279
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   106 AATAATGTCGTCTTATTTATTTAGCGGGTGAATATTTTATACTGTAAGTGAGCAATCAG 47

Qy  3280 AGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTTATATGTTT 3325
         |||||||||||||||||||||||||||||||||||||||||||| |
Db    46 AGTATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTTATATGTTT 1
```

RESULT 11
AI696721/c

LOCUS   AI696721   301 bp   mRNA   EST   03-JUN-1999
DEFINITION wc56d05.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2322633 3'
            similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   AI696721
NID   g4984621
VERSION   AI696721.1 GI:4984621
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 301)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Mar 10, 1998 this sequence version replaced gi:2948764.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Seq primer: -40UP from Gibco
            High quality sequence stop: 285.
FEATURES   Location/Qualifiers
     source   1..301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2322633"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   85 a   69 c   65 g   82 t
ORIGIN

Query Match   8.3%;   Score 282.6; DB 50; Length 301;
Best Local Similarity 98.6%;   Pred. No. 9.8e-40;
Matches 285; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   3043 ACCCTGTTTGGAGCTACTGCCAGAAGCACCAGAAGCACCAAAGTGCGGTTTCCCAAGCCTTTGTC 3102
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   289 AACCCTGTTGGAGCTACTGCCAGGACCAGAAGCACAAAGTGCGGTTTCCCAAGCCTTTGTC 230

Qy   3103 CATCTCAGCCCCCAGAGTATATCTGTCTTGGGGAATCCACACAGAAACTCAGGAGGACAC 3162
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   229 CATCTCAGCCCCCAGAGTATATCTGTCTTGGGGAATCCACACAGAAACTCAGGAGGACAC 170

Qy   3163 CCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT 3222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   169 CCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTTAAGTGCCGTTTGCAAT 110

Qy   3223 AATGTCGTCTTTATTTATTTATTAGCGGGTGAATATTTATACTGTAAGTGAGCAATCAGAGT 3282
          |||||||||||||||||||||||||||||||||||||||||||||

Db   109 AATGTCGTCTTTATTTATTTATTAGCGGGGTGAAATATTTATACTGTAAGTGAGCAATCAGAGT 50

Qy   3283 ATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAA 3331
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   49 ATAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTTAAAAA 1

RESULT 12
AI468280/c
LOCUS   AI468280   299 bp   mRNA   EST   30-MAR-1999
DEFINITION tg57a01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2112840 3'
            similar to contains element PTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AI468280
NID   g4330370
VERSION   AI468280.1 GI:4330370
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 299)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   On Mar 10, 1998 this sequence version replaced gi:2948787.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 427   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 286.
FEATURES   Location/Qualifiers
     source   1..299
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2112840"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                     with a modified polylinker; Plasmid DNA from the
                     normalized library NCI_CGAP_Pr22 was prepared, and ss
                     circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     985608-986759, 1101192-1101959, and 1217928-1220615)."
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT   88 a   69 c   66 g   76 t
ORIGIN

Query Match   8.3%;   Score 281.8; DB 47; Length 299;
Best Local Similarity 99.0%;   Pred. No. 1.3e-39;
Matches 294; Conservative   0; Mismatches   2; Indels   1; Gaps   1;

Qy   3029 CCACCAGTCCACAACCCTGTTGGAGCTACTGCCAGGACCAGAGCACAAAGTGCGGTTT 3088
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   296 CCCACAGTCCACAACCCTG-TTGGAGCTACTGCCAGGACCAGAAGCACAAAGTGCGGTTT 238

```
Qy   3089  CCCAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAG  3148
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237  CCCAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGGGAATCTCACACAG   178

Qy   3149  AAACTCAGGAGGACCACCCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTAA  3208
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177  AAACTCAGGAGGACCACCCCCTGCCTGAGCTAAGGAGGTCTTATCTCTCAGGGGGGTTAA   118

Qy   3209  GTGCCGTTTGCAATAATGTCGTCTATTATTATTAGCGGGGTGAATATTTATACTGTAAG  3268
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    117  GTGCCGTTTGCAATAATGTCGTCTATTATTATTAGCGGGGTGAATATTTATACTGTAAG    58

Qy   3269  TGAGCAATGACAGATAAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTT  3325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57  TGAGCAATGACAGATAAATGTTTATGGTGACAAAATTAAAGGCTTTCTTATATGTTT     1
```

RESULT 13
AA112573/c
LOCUS   AA112573   342 bp   mRNA   EST   23-DEC-1997
DEFINITION zm28c12.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
            IMAGE:526966 5', mRNA sequence.
ACCESSION   AA112573
NID   g1665120
VERSION   AA112573.1   GI:1665120
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 342)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
  TITLE   Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT   On May 9, 1995 this sequence version replaced gi:802278.

            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 307.
FEATURES   Location/Qualifiers
     source   1..342
                     /organism="Homo sapiens"
                     /db_xref="GDB:3918395"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:526966"
                     /clone_lib="Stratagene pancreas (#937208)"
                     /lab_host="SOLR cells (kanamycin resistant)"
                     /note="Organ: pancreas; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
                     Oligo dT. Pancreatic adenocarcinoma cell line. Average
                     insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT   67 a   100 c   115 g   59 t   1 others

ORIGIN

Query Match   8.2%;   Score 280.6;   DB 28;   Length 342;
Best Local Similarity 95.8%; Pred. No. 2.2e-39;
Matches 320; Conservative 0; Mismatches 10; Indels 4; Gaps 3;

```
Qy   1396  GTCCCACAGTGTGGCCGTGGTGACAGCTT--CAGCCGCCCCTCACCGGTTCACCTTCTCA  1453
           ||| ||||||||||||||||||||||| || | |||||||||||||||||||||||||||
Db    339  GTCAAACAGTGTGGCCGTGGTGACAAGTTTCAGCCGCCCCTCACCGGTTCACCTTCTCA   280

Qy   1454  GCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCACCACCGGGAGAAGCAGGTGTTC  1513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279  GCCCTGCAGATCCTGCCCTACACACTGGCCTCCCTCACCACCGGGAGAAGCAGGTGTTC   220

Qy   1514  CTGCCCAAATACCGAGGGGACACTGGAGGTGCTGAGCAGTGAGGACAGCCTGATGACCAGC  1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219  CTGCCCAAATACCGAGGGGACACTGGAGGTGCTGAGCAGTGAGGACAGCCTGATGACCAGC   160

Qy   1574  TTCC-TGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGACACGTGGGTGCTGGAGG  1632
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159  TTCCTTGCCAGGCCCTAAGCCTGGAGCTCCCTTCCCTAATGACACGTGGGTGCTGGAGG   100

Qy   1633  CAGTGGCCTGCTGCTCCCACCTCCACCCGCGCCTCTGCGGGGGCCTCTGCCTGTGATGTCTCCGT  1692
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     99  CAGTGGCCTGCTGCTCCCACCTCCACCCG-NCTCTGCGGGGGCCTCTGCCTGTGATGTCTCCGT    41

Qy   1693  ACGTGTGGTGTGGGTGAGCCCCACCGAGGCCAGG  1726
           |||||||||||||||||||||||||||||||| ||
Db     40  ACGTGTGGTGTGGGTGAGCCCCACCGAGGCCCGG     7
```

RESULT 14
AA492342/c
LOCUS   AA492342   264 bp   mRNA   EST   19-AUG-1997
DEFINITION ng8id12.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:941207, mRNA
            sequence.
ACCESSION   AA492342
NID   g2221904
VERSION   AA492342.1   GI:2221904
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 264)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1407518.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
            M.D., Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            DNA Sequencing by: Greg Lennon, Ph.D.
            Clone Distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Insert Length: 427   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 254.
FEATURES   Location/Qualifiers
     source   1..264
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:941207"
                     /clone_lib="NCI_CGAP_Pr6"
```

```
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="vector: pAMP10; mRNA made from prostatic
 intraepithelial neoplasia (low-grade), cDNA made by
 oligo-dT priming.  Non-directionally cloned.
 Size-selected on agarose gel, average insert size 600 bp.
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT    79 a    58 c    50 g    77 t
ORIGIN

Query Match          7.7%;      Score 262.4;  DB 34;  Length 264;
Best Local Similarity  99.6%;   Pred. No. 2.9e-36;
Matches 263;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy  3076 CAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGG 3135
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   264 CAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCCCAGAGTATATCTGTGCTTGGG 205

Qy  3136 GAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTC 3195
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   204 GAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAGCTAAGGGAGGTCTTATCTCTC 145

Qy  3196 AGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTATTATTTAGCGGGGTGAATAT 3255
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   144 AGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTATTATTTAGCGGGGTGAATAT 85

Qy  3256 TTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTC 3315
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    84 TTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATGGTGACAAAATTAAAGGCTTTC 25

Qy  3316 TTATATGTTTAAAAAAAAAAAAAA 3339
         |||||||||||||||||||
Db    24 TTATATGTTTAAAAAAAAAAAAAA 1


RESULT 15
AA579735
LOCUS       AA579735       288 bp      mRNA                  EST        03-SEP-1997
DEFINITION  nf39g07.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:916188, mRNA
            sequence.
ACCESSION   AA579735
NID         g2257919
VERSION     AA579735.1  GI:2357919
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 288)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Nov 29, 1993 this sequence version replaced gi:636080.

            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,
            M.D., Michael Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

FEATURES             Location/Qualifiers
     source          1..288
                     Seq primer: -40m13 fwd, ET from Amersham.
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:916188"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
 strand cDNA was primed with oligo(dT)17 on 50 ng of
 DNAse-treated, total cellular RNA obtained from
 5,000-10,000 microdissected preneoplastic cells
 histologically-determined to be prostatic intraepithelial
 neoplasia 2 (PIN2) cells.  Double-stranded cDNA was
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the
 cDNA with an adaptor-specific primer, and the resulting
 PCR product subcloned into pAMP10 by the UDG-cloning
 method (Life Technologies).  Average insert size is 600
 bp. NOTE: Not directionally cloned.  This library was
 constructed by David Krizman."

BASE COUNT    81 a    60 c    67 g    80 t
ORIGIN

Query Match          7.5%;      Score 255.6;  DB 35;  Length 288;
Best Local Similarity  96.5%;   Pred. No. 4.5e-35;
Matches 272;  Conservative  0;  Mismatches  9;  Indels  1;  Gaps  1;

Qy  3055 GCTACTGCAGGACCAGGAGCACAAAGTGCGGTTCCCAAGCCTTTGTCCATCTCAGCCCC 3114
         ||| ||||||||||||||||||||||||||||| |  ||||||||||||||||||||||||
Db     1 GCTCGACCAGGACCAGGAGCACAAAGTGCGGTCCTCCCCAGCCTTTGTCCATCTCAGCCCC 60

Qy  3115 CAGAGTATATCTGTCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAG 3174
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CAGAGTATATCTGTCTTGGGGAATCTCACACAGAAACTCAGGAGCACCCCCTGCCTGAG 120

Qy  3175 CTAAGGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTA 3234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTAAGGGAGGTCTTATCTCTCAGGGGGGGTTTAAGTGCCGTTTGCAATAATGTCGTCTTA 180

Qy  3235 TTTATTAGCGGGGTGAATATTTATACTGTAAGTGAGCAATCAGAGTATAATGTTTATG 3294
         ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   181 TTTATTAGCGGGGTGAATATTTATACTGTAAGTGAGC-ATCAGAGTATAATGTTTATG 239

Qy  3295 GTGACAAAATTAAAGGCTTTCCTTATATGTTTAAAAAAAAAA 3336
         ||||||||||||||||||||||||||||||||||||||||||
Db   240 GTGACAAAATTAAAGGCTTTCCTTATATGTTTAAAAAAAAAA 281


Search completed: September 25, 1999, 12:04:51
Job time: 8095 sec
```

Wed Sep 29 14:27:00 1999

This Page Blank (uspto)

QM of: US-09-030-606-111 to: A_Geneseq_36:*   out_format : pfs

Date: Sep 25, 1999 11:35 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09030606/runat.24091999_171616_29804/app_query.fasta.1
-DB=A_Geneseq_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-111
Query length: 1289
Database: A_Geneseq_36:*
Database sequences: 188963
Database length: 23686106
Search time (sec): 185.540000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
A_Geneseq_36:W61618	+	1402.00	2582.13	8.7e-137	273	! Clone HPWAE25 of TM4SF superfam
A_Geneseq_36:W58380	+	1258.00	2315.93	6.7e-122	241	! Human secreted protein AR415_4.
A_Geneseq_36:W59954	+	1258.00	2315.93	6.7e-122	241	! Amino acid sequence of the huma
A_Geneseq_36:W69306	+	1258.00	2315.93	6.7e-122	241	! Prostate tumour specific gene c
A_Geneseq_36:W71870	+	1258.00	2315.93	6.7e-122	241	! Amino acid encoded by prostate
A_Geneseq_36:W75060	+	1209.00	2225.25	7.7e-117	233	! Human secreted protein encoded
A_Geneseq_36:Y11883	+	486.00	890.35	3.8e-42	108	! Human 5' EST secreted protein S
A_Geneseq_36:Y13159	+	480.00	879.96	1.5e-41	101	! Human secreted protein encoded
A_Geneseq_36:Y12381	+	473.00	867.41	8.0e-41	97	! Human 5' EST secreted protein SE
A_Geneseq_36:W47275	+	453.00	830.01	9.5e-39	99	! Human HPK-1A C4.8 protein. Nucle
A_Geneseq_36:Y11862	+	452.00	829.36	1.1e-38	89	! Human 5' EST secreted protein SE
A_Geneseq_36:W86331	+	382.00	694.14	2.5e-31	140	! Kidney injury associated molecu
A_Geneseq_36:W61622	+	297.00	530.16	2.0e-22	238	! Clone HTPBA27 of TM4SF superfam
A_Geneseq_36:W05732	+	294.50	524.21	3.8e-22	267	! Human metastasis tumour suppres
A_Geneseq_36:W61624	+	288.50	513.71	1.3e-21	252	! Clone HHFEK40 of TM4SF superfam
A_Geneseq_36:W94494	+	285.50	506.94	3.1e-21	280	! Human CD53-like transmembrane p
A_Geneseq_36:W61620	+	282.00	501.95	7.2e-21	245	! Clone HSBBF02 of TM4SF superfam
A_Geneseq_36:W71708	+	281.50	500.13	8.4e-21	265	! Human integral membrane protein
A_Geneseq_36:R22360	+	277.50	493.97	3.3e-21	237	! CO-029 tumour associated antige
A_Geneseq_36:W61617	+	271.00	481.51	9.9e-20	245	! Clone HTEDK48 of TM4SF superfam
A_Geneseq_36:R20818	+	269.00	479.07	1.5e-19	219	! CD53 haematopoietic antigen. Ne
A_Geneseq_36:R91446	+	269.00	479.07	1.5e-19	219	! Human CD53 antigen. Cloning of
A_Geneseq_36:W89152	+	269.00	479.07	1.5e-19	219	! Human CD53 antigen. Cloning of
A_Geneseq_36:W80455	+	269.00	479.07	1.5e-19	219	! Human CD53 antigen. New cloning
A_Geneseq_36:W27333	+	266.00	471.85	3.3e-19	253	! Human membrane antigen TM4 supe
A_Geneseq_36:W61621	+	266.00	471.85	3.3e-19	253	! Clone HLTAH80 of TM4SF superfam
A_Geneseq_36:W74469	+	266.00	471.85	3.3e-19	253	! HP00966 protein sequence. Deter
A_Geneseq_36:W61623	+	253.50	450.16	6.1e-18	221	! Clone HAIDQ59 5' of TM4SF superf
A_Geneseq_36:Y07872	+	248.00	439.99	2.3e-17	220	! Human secreted protein fragment
A_Geneseq_36:W75128	+	227.00	419.47	1.6e-15	43	! Human secreted protein encoded b
A_Geneseq_36:R27525	+	226.00	398.70	4.3e-15	228	! Metastasis controlling peptide.
A_Geneseq_36:W70319	+	219.00	382.81	2.6e-14	294	! Secreted protein BD380_1. New l
A_Geneseq_36:R86834	+	217.00	382.02	3.7e-14	227	! Human CD9 sequence. Selectively
A_Geneseq_36:R33259	+	213.00	375.05	9e-14	218	! Sj23-like protein. Synthetic Sj
A_Geneseq_36:Y12214	+	147.00	258.79	5.0e-07	124	! Human 5' EST secreted protein S
A_Geneseq_36:W75129	+	140.50	250.99	2.0e-06	85	! Human secreted protein encoded b
A_Geneseq_36:W87503	+	129.50	200.42	9.0e-05	1212	! Human N-methyl-D-aspartate rec
A_Geneseq_36:W74838	+	125.50	213.19	0.0001	204	! Human N-methyl-D-aspartate rec
A_Geneseq_36:W92654	+	125.50	213.15	0.0001	204	! Human HT4P protein. New substan
A_Geneseq_36:W87504	+	125.00	193.56	0.0002	1061	! Human N-methyl-D-aspartate rec
A_Geneseq_36:Y04954	+	115.00	181.99	0.0020	572	! Mycobacterium species protein s
A_Geneseq_36:W81589	+	108.00	173.41	0.0091	387	! Protein encoded by human UCP3 g

A_Geneseq_36:Y11936   +   104.00   184.28   0.0113   77 ! Human 5' EST secreted protein
A_Geneseq_36:R95021   +   109.00   152.00   0.0181   3011 ! Hepatitis GB virus (HGBV) c
A_Geneseq_36:W30559   +   101.50   159.64   0.0456   449 ! Aspergillus oryzae hemA dele

seq_name: A_Geneseq_36:W61618

seq_documentation_block:
ID   W61618 standard; Protein; 273 AA.
AC   W61618;
DT   27-OCT-1998 (first entry)
DE   Clone HPWAE25 of TM4SF superfamily.
KW   Human; receptor; immune disorder; cancers; blood disorder;
KW   juvenile rheumatoid arthritis; Graves disease.
OS   Homo sapiens.
PN   WO9831799-A2.
PD   23-JUL-1998.
PF   21-JAN-1998; U00959.
PR   21-JAN-1997; US-034205.
PR   21-JAN-1997; US-034204.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gentz RL, Ni J, Rosen CA;
DR   WPI: 98-427555/36.
DR   N-PSDB; V48113.
PT   New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)
PT   - used to develop products for diagnosing or treating e.g. immune
PT   disorders, cancers, blood disorders or immuno-compromised disease
PT   states
PS   Claim 11; Page 36-37; 79pp; English.
CC   Clone HPWAE25 is a member of the TM4SF receptor superfamily. The
CC   products generated using the receptor can be used for treating abnormal
CC   conditions related to both an excess of and insufficient amounts of
CC   receptor activity. They can be used in the treatment of e.g. immune
CC   disorders, cancers, blood disorders, juvenile rheumatoid arthritis,
CC   Graves disease or immunocompromised disease states. The products can
CC   also be used for detection and diagnosis.
SQ   Sequence 273 AA;

alignment_scores:
     Quality: 1402.00   Length: 273
     Ratio: 5.154   Gaps: 0
Percent Similarity: 99.634   Percent Identity: 99.634

alignment_block:
US-09-030-606-111 x W61618

Align seg 1/1 to: W61618 from: 1 to: 273

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCT 172
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheLeuLysThrMetMetIleLeuPheAsnLeuLe 17

173 CATCTTCTGTGTGGTGCAGCCCGTGTGGCAGTGGGCATCGGGTGTCAA 222
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysValGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34

223 TCGATGGGGCATCCTTCTGAAGATCTTCGGGCCACTGTCCTCCCAGTGCC 272
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50

273 ATGCAGTTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGTGTGGT 322
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67

323 CTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84

373 GTGCCCCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTGAG 422
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 84 ysAlaLeuValThrPhePhePheIleLeuLeuIlePheIleAlaGlu 100

423 GTTGCAGCTGCTCGTGTGGTCGCCTTGGTGTACACCACCAAATGGCTGAGCACTT 472

```
CC     protein.  Its amino acid sequence was deduced from a full-length
CC     AT415_4 cDNA clone (see V30916) isolated from a human adult retina
CC     cDNA library.  The predicted amino acid sequence shows homology to
CC     human M35252 and CO-029 tumour associated antigens, and computer
CC     predictions suggest a potential transmembrane domain centered
CC     around amino acid 100 of the protein.  11 Novel human secreted
CC     proteins (see W58580-90) are claimed.  These can be expressed in
CC     recombinant host cells for analysis, characterisation, diagnostic
CC     or therapeutic use.  They can also be used as tissue or mol.wt.
CC     markers, to generate antibodies, and in interaction trap assays.
CC     They may have biological activities, e.g. cytokine, immunomodulator,
CC     haematopoiesis regulating activity, tissue growth activity, activin
CC     or inhibin activity, chemotactic or chemokinetic activity,
CC     haemostatic and thrombolytic activity, receptor/ligand activity,
CC     antiinflammatory, cadherin and tumour invasion suppressor activity,
CC     and tumour inhibition activity.  The proteins can be expressed in
CC     vivo from DNA, introduced in gene therapy vectors.
SQ     Sequence     241 AA;

alignment_scores:
       Quality: 1258.00              Length:    241
       Ratio:      5.220              Gaps:        0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-030-606-111 x W58380

Align seg 1/1  to: W58380    from: 1  to: 241

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGAATCCTCTTCCAATTTGCT 172
    |||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetIleLeuPheAsnLeuLeu   17

173 CATCTTTCGTGTGGTGCAGCCCTGTTGGCAGTGGCATCTGGGTGTCAA  222
    |||||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleLeuValSerI  34

223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCC 272
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla  50

273 ATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAGCCAGCCGTTGTGGT 322
    |||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaAlaGlyValValVa  67

323 CTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGCTGCTGAGAGCAAGT 372
    |||||||||||||||||||||||||||||||||||||||||||||||||
 67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC   84

373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCCTCCTCATTGCTGAG  422
    |||||||||||||||||||||||||||||||||||||||||||||||||
 84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuLeuePheIleAlaGlu 100

423 GTTCAGCTGCTGTGGTCGCCTTGGTGTACACCAACAATGGCTGAGCACTT 472
    |||||||||||||||||||||||||||||||||||||||||||||||||
101 ValAlaAlaAlaValValAlaLeuValTyrThrThrMetAlaGluHisPh  117

473 CCTGACGTTGCTGGTAGTGCCTGCCATCAAGAAGAAGAGATTATGGTTCCCAGG 522
    |||||||||||||||||||||||||||||||||||||||||||||||||
117 eLeuThrLeuLeuValValProAlaIleLysLysLysAspTyrGlySerGlnG 134

523 AAGACTTCACTCAAGTGTGGAACACCACCACCATGAAAGGGCTCAAGTGCTGT 572
    |||||||||||||||||||||||||||||||||||||||||||||||||
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys  150

573 GGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAA 622
    |||||||||||||||||||||||||||||||||||||||||||||||||
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs  167

623 CAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCAACACAGCCA  672
    |||||||||||||||||||||||||||||||||||||||||||||||||
167 nSerAlaPheProProPheCysAsnAspAsnValThrAsnThrAlaA  184
```

```
     |||||||||||||||||||||||||||||||||||||||||||||
101 ValAlaAlaAlaValValAlaLeuValTyrThrThrMetAlaGluHisPh 117

473 CCTGACGTTGCTGGTAGTGCCTGCCATCAAGAAGAAGAGATTATGGTTCCCAGG 522
    ||||||||||||||||||||||||||||||||||||||||||||
117 eLeuThrLeuLeuValValProAlaIleLysLysLysAspTyrGlySerGlnG 134

523 AAGACTTCACTCAAGTGTGGAACACCACCACCATGAAAGGGCTCAAGTGCTGT 572
    ||||||||||||||||||||||||||||||||||||||||||||
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys 150

573 GGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAA 622
    ||||||||||||||||||||||||||||||||||||||||||||
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167

623 CAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCAACACAGCCA 672
    ||||||||||||||||||||||||||||||||||||||||||||
167 nSerAlaPheProProPheCysAsnAspAsnValThrAsnThrAlaA 184

673 ATGAAACCTGCACCAAGCAAAAGGCTCACGACCCAAAAAGTAGAGGGTTGC 722
    ||||||||||||||||||||||||||||||||||||||||||||
184 snGluThrCysThrLysSerGlnLysAlaHisAspGlnLysValGluGlyCys 200

723 TTCAATCAGCTTTTGATGACATCCGAACTAATGCAGTCACCGTGGGTGG 772
    ||||||||||||||||||||||||||||||||||||||||||||
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyGl 217

773 TGTGGCAGCTGGAATTGGGGGCCTCGAGCTGGCTGCCATGAT.TGTGTCC 821
    ||||||||||||||||||||||||||||||||||        ||||||
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetAsnCysValH 234

822 ATGTATCTGTACTGCAATCTACAATAAGTCCACTTCTGCCCTGCCACTA 871
    ||||||||||||||||||||||||||||||||||||||||||||
234 isValSerValLeuGlnSerThrIleSerProLeuLeuProLeu 250

872 CTGCTGCCACATGGAACTGTGAAGAGGCACCCTGGCAAGCAGCAGTGAT 921
    ||||||||||||||||||||||||||||||||||||||||||||
251 LeuLeuProHisGlyAsnCysGluGluAlaProTrpGlnAlaAlaValIl 267

922 TGGGGAGGGACAGGATC 940
    |||||||||||||
267 eGlyGlyGlyAspArgIle 273
```

seq_name: A_Geneseq_36:W58380

```
seq_documentation_block:
ID    W58380 standard; Protein; 241 AA.
AC    W58380;
DT    14-SEP-1998  (first entry)
DE    Human secreted protein AR415_4.
KW    AR415_4; secreted protein; protein factor; human.
OS    Homo sapiens.
FH                    Location/Qualifiers
FT    Key
FT    Peptide         14..26
FT                    /note= "putative leader/signal peptide or
FT                    transmembrane domain"
FT    Protein         27..241
FT                    /label= Mat_protein
PN    WO9817687-A2.
PD    30-APR-1998.
PF    24-OCT-1997; U19590.
PR    24-OCT-1997; US-740274.
PR    25-OCT-1996; US-740274.
PA    (GEMY) GENETICS INST INC.
PI    Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI    Racie LA, Spaulding V, Treacy M;
DR    WPI: 98-261426/23.
DR    N-PSDB; V30916.
PT    Nucleic acid encoding secreted protein from human cells - useful,
PT    e.g. as immunomodulator, antitumour agent, promoters of tissue
PT    growth, haemostatic and thrombolytic agents etc.
PS    Claim 8; Page 67-68; 114pp; English.
CC    This polypeptide, designated AT415_4, is a novel human secreted
```

```
123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCT  172
        ||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetIleLeuPheAsnLeuLe   17

173 CATCTTTCGTGTGGTGCAGCCCTGGTTGGCAGTGGGCATCGGTGGTCAA  222
        ||||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI  34

223 TCGATGGGCACCTTTCTGAAGACATTCGGGCCACTGTCGTCCAGTGCC   272
        |||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla    50

273 ATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT  322
        ||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa  67

323 CTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGACTGCTGAGAGCAAGT  372
        ||||||||||||||||||||||||||||||||||||||||
 67 lPheAlaLeuGlyPheLeuGlyLysTyrGlyAlaLysThrGluSerLysC  84

373 GTGCCCTCGTGACGTTCTTCTCATCCTCCCTCCATCTTCATTGCTGAG  422
        ||||||||||||||||||||||||||||||||||||||||||||||||
 84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAlaGlu 100

423 GTTGCAGCTGCTGTGGTGCGCCTTGGTGTACACCACAATGGCTGAGCACTT  472
        ||||||||||||||||||||||||||||||||||||||||||||||
101 ValAlaAlaAlaValAlaLeuValTyrThrThrMetAlaGluHisPh    117

473 CCTGAGCGTTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGG  522
        |||||||||||||||||||||||||||||||||||||||||||||
117 eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG 134

523 AAGACTTCACTCAAGTGTGGAACACCACCACCATGAAAGGGCTCAAGTGCTGT  572
        |||||||||||||||||||||||||||||||||||||||
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys  150

573 GGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAA  622
        ||||||||||||||||||||||||||||||||||||||||||||||||
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs 167

623 CAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCAACACAGCCA  672
        ||||||||||||||||||||||||||||||||||||||||||||||||
167 nSerAlaPheProProPheCysAsnAspAsnValThrAsnThrAlaA   184

673 ATGAAAACCTGCCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTGC  722
        ||||||||||||||||||||||||||||||||||||||||||
184 snGluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys  200

723 TTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGG  772
        |||||||||||||||||||||||||||||||||||||||||
201 PheAsnGlnLeuCysTyrAspIleArgThrAsnAlaValThrValGlyGl  217

773 TGTGGCAGCTGGAATTGGGGGCCTCGAGCTGCCATGATTGTGTCCA  822
        ||||||||||||||||||||||||||||||||||||||||||||||
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM 234

823 TGTATCTGTACTGCAATCTACAA  845
        |||||||||||||||||||||||
234 etTyrLeuTyrCysAsnLeuGln 241
```

seq_name: A_Geneseq_36:W69386

seq_documentation_block:
ID   W69386 standard; Protein; 241 AA.
AC   W69386;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone N1-1862 protein.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.

```
673 ATGAAAACCTGCCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTGC  722
        ||||||||||||||||||||||||||||||||||||||||||
184 snGluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys  200

723 TTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGG  772
        |||||||||||||||||||||||||||||||||||||||||
201 PheAsnGlnLeuCysTyrAspIleArgThrAsnAlaValThrValGlyGl  217

773 TGTGGCAGCTGGAATTGGGGGCCTCGAGCTGCCATGATTGTGTCCA  822
        ||||||||||||||||||||||||||||||||||||||||||||||
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM 234

823 TGTATCTGTACTGCAATCTACAA  845
        |||||||||||||||||||||||
234 etTyrLeuTyrCysAsnLeuGln 241
```

seq_name: A_Geneseq_36:W59954

seq_documentation_block:
ID   W59954 standard; Protein; 241 AA.
AC   W59954;
DT   02-DEC-1998 (first entry)
DE   Amino acid sequence of the human tumour-associated antigen.
KW   Human; tumour-associated antigen; PRAT; stimulation; cell proliferation;
KW   antagonist; cancer; genetic defect; sickle cell anaemia; agonist;
KW   antibody; hybridisation; probe.
OS   Homo sapiens.
PN   WO983810-A1.
PD   03-SEP-1998.
PF   27-FEB-1998; U03953.
PR   28-FEB-1997; US-808148.
PA   (INCY-) INCYTE PHARM INC.
PI   Goli SK, Hillman JL;
DR   WPI; 98-481208/41.
DR   N-PSDB; V54014.
PT   Human tumour-associated antigen PRAT - useful for stimulating cell
PT   proliferation and screening for antagonists useful to treat or
PT   prevent cell proliferation disorders e.g. cancers
PS   Disclosure: Fig 1A-1C; 54pp; English.
CC   This is the amino acid sequence of the human tumour-associated antigen
CC   (PRAT) used in the method of the invention for stimulating cell
CC   proliferation and screening for antagonists useful to treat or prevent
CC   cell proliferation disorders such as cancer, and genetic defect e.g.
CC   sickle cell anaemia. The polypeptides can be combined with a suitable
CC   carrier in pharmaceutical compositions, and also used to screen for
CC   antagonists, agonists, and to generate antibodies. PRAT agonists can
CC   be added to a cell to stimulate cell proliferation by increasing or
CC   prolonging the activity of PRAT as above. The antagonists can be
CC   combined with a suitable carrier in pharmaceutical compositions, which
CC   can be administered to subjects to treat or prevent disorders
CC   associated with cell proliferation, especially cancers. Antibodies
CC   specific for PRAT may be used directly as antagonists, or indirectly
CC   as a targeting or delivery mechanism to bring pharmaceutical agents to
CC   PRAT-expressing cells. They are also useful to diagnose conditions or
CC   diseases characterised by PRAT expression and to monitor therapeutic
CC   interventions. The polynucleotide encoding PRAT, or complementary
CC   sequences, can be used to produce hybridisation probes, useful to
CC   detect polynucleotides or fragments encoding PRAT, e.g. to diagnose
CC   diseases relating to polypeptide expression or monitor PRAT regulation
CC   during therapeutic intervention.
SQ   Sequence 241 AA;

alignment_scores:
       Quality: 1258.00   Length: 241
       Ratio: 5.220   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-030-606-111 x W59954

Align seg 1/1 to: W59954   from: 1 to: 241

```
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
     WPI; 98-480805/41.
DR   N-PSDB: V58587.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Example 1; Page 89-90; 141pp; English.
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence   241 AA;
```

```
alignment_scores:
     Quality: 1258.00            Length:      241
     Ratio:   5.220              Gaps:          0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-09-030-606-111 x W69386

Align seg 1/1  to: W69386  from: 1  to: 241

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTTCTTCAATTTGCT 172
    |||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetIleLeuPheAsnLeuLe    17

173 CATCTTTCTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGTGTCAA    222
    ||||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI  34

223 TCGATGGGGCATCCTTTCTGAAGAGATCTCGGGCCACTGTCGTCCAGTGCC 272
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla     50

273 ATGCAGTTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT   322
    |||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa  67

323 CTTTGCCTCTGGTTCCCGGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT  372
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC  84

373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCAATCTTCATTGCTGAG 422
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuLeuIlePheIleAlaGlu 100

423 GTTGCAGCTGCTGTGGTCGCCTTGGTGTACACCACCAATGGCTGAGCACTT 472
    |||||||||||||||||||||||||||||||||||||||||||||||||||
101 ValAlaAlaAlaAlaValValAlaLeuValThrThrMetAlaGluHisPh  117

473 CCTGAGCGTGCTGGTAGTGCCTGCCATCCAAGAAAGAATTATGGTTCCCAGG 522
    ||||||||||||||||||||||||||||||||||||||||||||||||||
117 eLeuThrLeuLeuValValProAlaIleLysLysLysAspTyrGlySerGlnG 134

523 AAGACTTCACTCAAGTGTGGAACACCACCATGAAAGGGCTCAAGTGCTGT   572
    |||||||||||||||||||||||||||||||||||||||||||||||||
134 luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys  150

573 GGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAA   622
    |||||||||||||||||||||||||||||||||||||||||||||||||
151 GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs  167
```

```
623 CAGTGCCTTCCCCCATCTGTGTGCAATGACAAACGTCACCACCACCAGCCA 672
    |||||||||||||||||||||||||||||||||||||||||||||||||||
167 nSerAlaPheProProPheCysCysAsnAspAsnValThrAsnThrAlaA  184

673 ATGAAAACTGCCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTGC 722
    |||||||||||||||||||||||||||||||||||||||||||||||||||
184 snGluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys  200

723 TTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGG 772
    |||||||||||||||||||||||||||||||||||||||||||||||||||
201 PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyGl  217

773 TGTGGCAGCTGGAATTGGGGGCCTGAGCTGGCTGCCATGATTGTGTCCA   822
    |||||||||||||||||||||||||||||||||||||||||||||||||
217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM  234

823 TGTGATCGTGACTGCAATCTACAA 845
    ||||||||||||||||||||||||
234 etTyrLeuTyrCysAsnLeuGln 241
```

seq_name: A_Geneseq_36:W71870

```
seq_documentation_block:
ID   W71870 standard; Protein; 241 AA.
AC   W71870;
DT   06-JAN-1999  (first entry)
DE   Amino acid encoded by prostate tumour clone N1-1862.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
DR   N-PSDB; V61202.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Example 1; Page 84-85; 130pp; English.
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
SQ   Sequence   241 AA;
```

```
alignment_scores:
     Quality: 1258.00            Length:      241
     Ratio:   5.220              Gaps:          0
Percent Similarity: 100.000     Percent Identity: 100.000

alignment_block:
US-09-030-606-111 x W71870

Align seg 1/1  to: W71870  from: 1  to: 241  ..

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTTCTTCAATTGCT  172
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetIleLeuPheAsnLeuLeuLe  17

173 CATCTTTCTGTGGTGGTGCAGCCCTGTTGGCAGTGGGCATCGGGTGTCAA  222
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI  34

223 TCGATGGGGCATCCTTTCTGAAGATCTCGGGCCACTGTCGTCCAGTGCC   272
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerAla    50
```

```
273  ATGCAGTTGTCAACGTGGGCTACTTCCCATCGCAGCCGGCGGTTGTGGT  322
     ||||||||||||||||||||||||||||||||||||||||||||||||
 51  MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValva   67

323  CTTTGCTCCTTGGTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT  372
     ||||||||||||||||||||||||||||||||||||||||||||||||
 67  lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC   84

373  GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAG  422
     ||||||||||||||||||||||||||||||||||||||||||||||||
 84  ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePheIleAlaGlu  100

423  GTTGCAGCTGCTGCTGGTGGTCGCCTTGGTGTACACCACAATGGCTGAGCACTT  472
     ||||||||||||||||||||||||||||||||||||||||||||||||
101  ValAlaAlaAlaValAlaLeuValTyrThrThrMetAlaGluHisPh  117

473  CCTGACGTTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGG  522
     ||||||||||||||||||||||||||||||||||||||||||||||||
117  eLeuThrLeuLeuValValProAlaIleLysLysAspTyrGlySerGlnG  134

523  AAGACTTCACTCAAGTGTGGAACACCACCATGAAAGGGCTCAAGTGCTGT  572
     ||||||||||||||||||||||||||||||||||||||||||||||||
134  luAspPheThrGlnValTrpAsnThrThrMetLysGlyLeuLysCysCys  150

573  GGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAA  622
     ||||||||||||||||||||||||||||||||||||||||||||||||
151  GlyPheThrAsnTyrThrAspPheGluAspSerProTyrPheLysGluAs  167

623  CAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCACCAACACCAGCCA  672
     ||||||||||||||||||||||||||||||||||||||||||||||||
167  nSerAlaPheProProPheCysCysAsnAspAsnValThrAsnThrAlaA  184

673  ATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTGC  722
     ||||||||||||||||||||||||||||||||||||||||||||||||
184  snGluThrCysThrLysGlnLysAlaHisAspGlnLysValGluGlyCys  200

723  TTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTGGGTGG  772
     ||||||||||||||||||||||||||||||||||||||||||||||||
201  PheAsnGlnLeuLeuTyrAspIleArgThrAsnAlaValThrValGlyGl  217

773  TGTGGCAGCTGGAATTGGGGGCCTCCAGCTGGCTGCCATGATTGTGTCCA  822
     ||||||||||||||||||||||||||||||||||||||||||||||||
217  yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSerM  234

823  TGTATCTGTACTGCAATCTACAA  845
     |||||||||||||||||||||||
234  etTyrLeuTyrCysAsnLeuGln  241
```

seq_name: A_Geneseq_36:W75060

seq_documentation_block:
ID   W75060 standard; Protein; 233 AA.
AC   W75060;
DT   28-JAN-1999 (first entry)
DE   Human secreted protein encoded by gene 4 clone HKCSR70.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue: cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   Misc-difference 233
FT   /label= unknown
PN   WO9839446-A2.
PD   11-SEP-1998.
PF   06-MAR-1998; U04492.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043315.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043669.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   23-MAY-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047581.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   22-AUG-1997; US-056630.
PR   22-AUG-1997; US-056631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056662.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-058862.
PR   22-AUG-1997; US-058864.
PR   22-AUG-1997; US-058872.
PR   22-AUG-1997; US-058874.
PR   22-AUG-1997; US-058875.
PR   22-AUG-1997; US-058876.
PR   22-AUG-1997; US-058877.
PR   22-AUG-1997; US-058878.
PR   22-AUG-1997; US-058879.
PR   22-AUG-1997; US-058880.

```
PR    22-AUG-1997; US-056881.
PR    22-AUG-1997; US-056882.
PR    22-AUG-1997; US-056884.
PR    22-AUG-1997; US-056886.
PR    22-AUG-1997; US-056887.
PR    22-AUG-1997; US-056888.
PR    22-AUG-1997; US-056889.
PR    22-AUG-1997; US-056892.
PR    22-AUG-1997; US-056893.
PR    22-AUG-1997; US-056894.
PR    22-AUG-1997; US-056903.
PR    22-AUG-1997; US-056908.
PR    22-AUG-1997; US-056909.
PR    22-AUG-1997; US-056910.
PR    22-AUG-1997; US-056911.
PR    05-SEP-1997; US-057650.
PR    05-SEP-1997; US-057761.
PA    (HUMA-) HUMAN GENOME SCI INC.
PI    Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI    Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,
PI    Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI    Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR    WPI; 98-609887/51.
DR    N-PSDB; V34157.
PT    New isolated human genes and the secreted polypeptides they encode
PT    - useful for diagnosis and treatment of e.g. cancers, neurological
PT    disorders, immune diseases, inflammation or blood disorders
PS    Claim 1: Page 281-282; 447pp; English.
CC    This sequence represents a secreted human protein encoded by the gene
CC    clone detailed in the descriptor line.
CC    The gene can be used to generate fusion proteins by linking to the gene
CC    to a human immunoglobulin Fc portion (e.g. V34145) for increasing the
CC    stability of the fused protein as compared to the human protein only.
CC    The invention relates to 70 novel genes and their fragments (nucleic acid
CC    sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC    are useful for preventing, treating or ameliorating medical conditions
CC    e.g. by protein or gene therapy. Also, pathological conditions can be
CC    diagnosed by determining the presence of mutations in the new polynucleotides.
CC    Specific uses are described for each of the 70 polynucleotides, based on
CC    which tissues they are most highly expressed in (see V34154 for described
CC    or by determining the amount of the new polypeptides in a sample
CC    uses).
SQ    Sequence   233 AA;
```

```
alignment_scores:
     Quality:   1209.00          Length:     232
     Ratio:        5.211          Gaps:          0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-030-606-111 x W75060

Align seg 1/1  to: W75060   from: 1  to: 233
```

```
150 ATGATGATCCTCTTCAATTGCTCATCTTTCTGTGTGTGCAGCCCTGTT 199
    ||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetMetIleLeuPheAsnLeuLeuIlePheLeuCysSerGlyAlaAlaLeuLeu 17

200 GGCAGTGGCAATCGGGTCTCAATCGATGGGCCATCCTTTCTGAAGATCT 249
    ||||||||||||||||||||||||||||||||||||||||||||||||
 17 uAlaValGlyIleTrpValSerIleAspIleAlaSerPheLeuSerIleP 34

250 TCGGGCCACTGTCCTCCAGTGCCATGCAGTTTGTCAACGTGGGCTACTTC 299
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 heGlyProLeuSerSerAlaMetGlnPheValAsnValGlyTyrPhe 50

300 CTCATCAGCAGCCGGCGGTGTGGTCTTTGCTCTTGGTTTCCTGGGCTGCTA 349
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LeuIleAlaAlaGlyValValValPheAlaLeuGlyPheLeuGlyCysTy 67

350 TGGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGGACGTTCTTCTTCATCC 399
    ||||||||||||||||||||||||||||||||||||||||||||||||||

 67 rGlyAlaLysThrGluSerLysCysAlaLeuValThrPhePheIleL 84

400 TCCTCCTCATCTTCATTGCTGAGGTTGCAGCTGCTGGTCGCCTTGGTG 449
    |||||||||||||||||||||||||||||||||||||||||||||||
 84 euLeuLeuIlePheIleAlaGluValAlaAlaAlaValAlaLeuVal 100

450 TACACCACAATGGCTGAGCACTTCTGACGTTGCTGGTAGTGCCTGCCAT 499
    |||||||||||||||||||||||||||||||||||||||||||||||
101 TyrThrThrMetAlaGluHisPheLeuThrLeuLeuValValProAlaIl 117

500 CAAGAAAGATATGGTTCCCAGGAGACTTCACTCAAGTGTGGAACACCA 549
    ||||||||||||||||||||||||||||||||||||||||||||||||
117 eLysLysAspTyrGlySerGlnGluAspPheThrGlnValTrpAsnThrT 134

550 CCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAG 599
    ||||||||||||||||||||||||||||||||||||||||||||||||||
134 hrMetLysGlyLeuLysCysCysGlyPheThrAsnTyrThrAspPheGlu 150

600 GACTCACCCTACTTCAAAGAGAACAAGTGCCTTTCCCCATTCTGTTGCAA 649
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 AspSerProTyrPheLysGluAsnSerAlaPheProProPheCysCysAs 167

650 TGACACGTCACCAACACAGCCAATGAAACCTGCACCAAGCAAAGGCTC 699
    ||||||||||||||||||||||||||||||||||||||||||||||||
167 nAspAsnValThrAsnThrAlaAsnGluThrCysThrLysGlnLysAlaH 184

700 ACGACCAAAAAGTAGAGGGTTGCTTCAATCAGCTTTTGATGACATCCGA 749
    |||||||||||||||||||||||||||||||||||||||||||||||||
184 isAspGlnLysValGluGluTyrCysPheAsnGlnLeuLeuTyrAspIleArg 200

750 ACTAATGCAGTCACCGTGGGTGGTGTGGCAGCTGGAATTGGGGGCCTGA 799
    ||||||||||||||||||||||||||||||||||||||||||||||||
201 ThrAsnAlaValThrValGlyGlyValAlaAlaGlyIleGlyGlyLeuGl 217

800 GCTGGCTGCCATGATTGTGTCCCATGTATCGTACTGCAATCTACAA 845
    ||||||||||||||||||||||||||||||||||||||
217 uLeuAlaAlaMetIleValSerMetTyrLeuTyrCysAsnLeuGln 232
```

```
seq_name: A_Geneseq_36:Y11883

seq_documentation_block:
ID   Y11883 standard; Protein; 108 AA.
AC   Y11883;
DT   18-JUN-1999  (first entry)
DE   Human 5' EST secreted protein SEQ ID NO: 483.
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition activity
OS   Homo sapiens.
PN   WO9906550-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1232.
PR   01-AUG-1997; US-905144.
PA   (GEST ) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153780/13.
DR   N-PSDB; X40605.
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
PS   Claim 34; Page 605; 675pp; English.
CC   X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins expressed in prostate, and encode the proteins given in
CC   Y11716 to Y11993 respectively. The proteins given represent the signal
CC   peptide and an N-terminal fragment of a secreted protein, the nucleic
CC   acid sequences can be used for producing secreted human gene products.
CC   They can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell proliferation and
CC   differentiation activity, haematopoiesis regulating activity, tissue
CC   growth regulating activity, reproductive hormone regulating activity,
```

CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC activity or other activities. The products can be used in forensic, gene
CC therapy and chromosome mapping procedures. The sequences can also be used
CC for obtaining corresponding promoter sequences. The nucleic acids
CC encoding the signal peptides can be used for directing extracellular
CC secretion of a polypeptide or the insertion of a polypeptide into a
CC membrane, or importing a polypeptide into a cell.
SQ Sequence   108 AA;

alignment_scores:
        Quality: 486.00   Length: 109
        Ratio: 4.542   Gaps: 1
  Percent Similarity: 98.165   Percent Identity: 93.578

alignment_block:
  US-09-030-606-111 x Y11883

Align seg 1/1 to: Y11883   from: 1   to: 108

273 ATGCAGTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT 322
    ||||||| ||:::||||||||||||||||||||||||||||||||||||
  1 MetGlnPhe.***ThrTrpAlaThrSerSerSerGlnProAlaLeuTrpS 17

323 CTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGC..AAG 371
    ||||||||||||||||||||||||||||||||||||||||| ||| |||
 17 erLeuLeuLeuValSerTrpAlaAlaMetValLeuArgLeuArgSerLys 33

372 TGTGCCCCTGCTGAGGTTCTTCTTCATCCCTCCTCCTCCTCATCCATGCTGA 421
    |||||||||||||||||||||||||||||||||||||||||||||||||:||
 34 CysAlaLeuValThrPhePhePheLeuLeuLeuLeuIlePheIleAlaGlI 50

422 GGTTGCAGCTGCTGCTGGTGCCCCTTGGTGTACCACCAAATGGCTGAGCACT 471
    ||||||||||||||||||||:::|||||::::||||||||||||||||||||
 50 uValAlaAlaAlaValAlaValAlaLeuValTyr***ThrMet***GluHisP 67

472 TCCTGACGTTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAG 521
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 heLeuThrLeuLeuValProAlaIleLysLysAspTyrGlySerGln 83

522 GAAGACTTCACTCAAGTGTGGAACACCACCATGAAAGGGCTCAAGTGCTG 571
    ||||||||||||:::||||*||||||||||||||||||||||||||||||
 84 GluAspPheThrGlnVal***AsnThrThrMetLysGlyLeuLysCysCy 100

572 TGGCTTCACCAACTATACGGATTTT 596
    |||||||||||||||||::::
100 sGlyPheThrAsnTyrThrAspTrp 108

seq_name: A_Geneseq_36:Y13159

seq_documentation_block:
ID Y13159 standard; Protein; 101 AA.
AC Y13159;
DT 22-JUN-1999 (first entry)
DE Human secreted protein encoded by 5' EST SEQ ID NO: 173.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN WO9905552-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1236.
PR 01-AUG-1997; US-905223.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153782/13.
DR N-PSDB; X51959.
PT New isolated brain-derived nucleic acids - used to develop products
PT which may have cytokine, immune, regulatory, haematopoiesis

PT regulating, anti-inflammatory or tumour inhibition activity
PS Claim 34; Page 543-544; 577pp; English.
CC X51787 to x52019 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12987 to Y13219,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence   101 AA;

alignment_scores:
        Quality: 480.00   Length: 101
        Ratio: 4.898   Gaps: 0
  Percent Similarity: 97.030   Percent Identity: 95.050

alignment_block:
  US-09-030-606-111 x Y13159

Align seg 1/1 to: Y13159   from: 1   to: 101

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCT 172
    |||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17

173 CATCTTTCTGTGTGGTGCAGCCCTGTTGGGCAGTGGGCATCTGGGTGTCAA 222
    |||||||||||||||||||||||||||::::||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeu***ValGlyIleTrpValSerI 34

223 TCGATGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCC 272
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50

273 ATGCAGTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT 322
    |||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaLeuGlyValValVa 67

323 CTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372
    |||||||||||||||||||||| ||||||||||||||||||||||||||
 67 lPheAlaLeuGlyPheLeuGlyCysTyr***AlaLysThrGluSer***C 84

373 GTGCCCCTGCTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAG 422
    ||||||||||||||||||||||||| ||||||||||||||||||:::
 84 ysAlaLeuValThrProThr***IleLeuLeuLeuIlePheIleAlaAsp 100

423 GTT 425
    |||
101 Val 101

seq_name: A_Geneseq_36:Y12381

seq_documentation_block:
ID Y12381 standard; Protein; 97 AA.
AC Y12381;
DT 17-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO:412.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.

OS   Homo sapiens.
PN   WO9906548-A2.
PD   11-FEB-1999.
PF   31-JUL-1998; IB1222.
PR   01-AUG-1997; US-905135.
PA   (GEST) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153778/13.
DR   N-PSDB; X41214.
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT   kidney, lung, umbilical cord, placenta and colon tissue
PS   Claim 27; Page 732; 824pp; English.
CC   X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC   respectively. The proteins given represent the signal peptide and an
CC   N-terminal fragment of a secreted protein. The nucleic acid sequences
CC   can be used for producing secreted human gene products. They can also
CC   be used to develop products for diagnosis and therapy. The proteins
CC   obtained may have cytokine activity, cell proliferation/differentiation
CC   activity, haematopoiesis regulating activity, tissue growth regulating
CC   activity, reproductive hormone regulating activity, chemotactic/
CC   chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC   ligand activity, anti-inflammatory activity, tumour inhibition activity
CC   or other activities. The products can be used in forensic, gene therapy
CC   and chromosome mapping procedures. The sequences can also be used for
CC   obtaining corresponding promoter sequences. The nucleic acids encoding
CC   the signal peptide can be used for directing extracellular secretion of
CC   a polypeptide or the insertion of a polypeptide into a membrane, or
CC   importing a polypeptide into a cell.
SQ   Sequence   97 AA;

alignment_scores:
        Quality: 473.00   Length: 97
        Ratio: 4.927   Gaps: 0
Percent Similarity: 98.969   Percent Identity: 96.907

alignment_block:
US-09-030-606-111 x Y12381   ..

Align seg 1/1 to: Y12381 from: 1 to: 97

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCT 172
    ||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17

173 CATCTTTCTGTGTGGTGCAGCCCCTGTTGGCAGTGGGCATCTGGTGTCAA 222
    |||||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34

223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCC 272
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50

273 ATGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT 322
    |||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67

323 CTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372
    |||||||||||||||||||||||||||:|||:::::|||
 67 lPheAlaLeuGlyPheLeuGlyCys***GlyAlaLys***Glu***LysC 84

373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCCTCATCTTC 413
    ||||||||||||||||||||||||||||||||||||||
 84 ysAlaLeuValThrPhePhePheIleLeuLeuLeuIlePhe 97

seq_name: A_Geneseq_36:W47275

seq_documentation_block:
ID   W47275 standard; Protein; 99 AA.
AC   W47275;
DT   02-JUL-1998 (first entry)

DE   Human HPK-1A C4.8 protein.
KW   Cervical cancer; treatment; diagnosis; passage cell; lesion;
KW   human foreskin keratinocyte cell line; HPK-1A; antibody; smear.
OS   Homo sapiens.
PN   DE1964207-C1.
PD   26-FEB-1998.
PF   27-NOV-1996; 049207.
PR   27-NOV-1996; DE-049207.
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PI   Duerst M, Nees M;
DR   WPI; 98-121623/12.
DR   N-PSDB; V15558.
PT   Nucleic acid characteristic of late or early passage cells
PT   immortalised by papilloma virus - and related polypeptide(s) and
PT   antibodies, used for diagnosis and treatment of cervical cancer and
PT   assessing potential for progression of cervical lesions
PS   Claim 2; Fig 1; 8pp; German.
CC   This protein, C4.8, is derived from a human papillomavirus (HPV)
CC   immortalised human foreskin keratinocyte cell line HPK-1A and is
CC   characteristic of late or early passage cells. This sequence is used
CC   in a method for assessing the potential for progression of cervical
CC   lesions. Antibodies generated against the encoded polypeptide are used
CC   for diagnosis of cervical cancer and to assess potential for lesion
CC   progression. Antibodies can also be used therapeutically by inhibiting
CC   the polypeptide. Antisense molecules based on the nucleotide sequence
CC   are used to inhibit expression of the protein. Detecting polypeptides,
CC   or related RNA, characteristic of late passage cells (which are
CC   potentially malignant) in cervical smears is a reliable way of
CC   assessing progression potential.
SQ   Sequence   99 AA;

alignment_scores:
        Quality: 453.00   Length: 96
        Ratio: 4.819   Gaps: 0
Percent Similarity: 97.917   Percent Identity: 96.875

alignment_block:
US-09-030-606-111 x W47275   ..

Align seg 1/1 to: W47275 from: 1 to: 99

219 TCAATCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAG 268
    :::||||||||||||||||||||||||||||||||||||||||||||||
  1 AlaIleAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSe 17

269 TGCCATGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTG 318
    |||||||||||||||||||||||||||||||||||||||||||||||||
 17 rAlaMetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValV 34

319 TGGTCTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGC 368
    |||||||||||||||||||||||||||||||||||||||||||||||||
 34 alValPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSer 50

369 AAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCCTCATCTTCATTGC 418
    |||||||||||||||||||||||||||||||||||||||||||||||
 51 LysCysAlaLeuValThrPhePhePheIleLeuLeuLeuLeuIlePheAl 67

419 TGAGGTTGCAGCTGCGTGGTCGCCCTTGGTGTACACCACAATGGCTGAGC 468
    |||||||||||||||||||||||||||||| ||||||||||
 67 aGluValAlaAlaAlaValAlaAlaLeuValTyrThrIleMetAlaGluH 84

469 ACTTCCTGACGTTGCTCGTGAGTGCCTGCCATCAAGAAA 506
    |||| |||||||||||||||||||||||||||||||||
 84 isPheProThrLeuLeuValValProAlaIleLysLys 96

seq_name: A_Geneseq_36:Y11862

seq_documentation_block:
ID   Y11862 standard; Protein; 89 AA.
AC   Y11862;
DT   18-JUN-1999 (first entry)
DE   Human 5' EST secreted protein SEQ ID NO: 462.

KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
OS   Homo sapiens.
PN   WO9905550-A2.
PD   11-FEB-1999.
PF   31-JUL-1998;   IB1232.
PR   01-AUG-1997; US-905144.
PA   (GEST) GENSET.
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
DR   WPI; 99-153780/13.
DR   N-PSDB; X40584.
PT   New isolated prostate-derived nucleic acids - used to develop
PT   products which may have cytokine, immune regulatory, haematopoiesis
PT   regulating, anti-inflammatory or tumour inhibition activity
PS   Claim 34; Page 590-591; 675pp; English.
CC   X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC   secreted proteins expressed in prostate, and encode the proteins given in
CC   Y11716 to Y11993 respectively. The proteins given represent the signal
CC   peptide and an N-terminal fragment of a secreted protein. The nucleic
CC   acid sequences can be used for producing secreted human gene products.
CC   They can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell proliferation and
CC   differentiation activity, haematopoiesis regulating activity, tissue
CC   growth regulating activity, reproductive hormone regulating activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC   receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC   activity or other activities. The products can be used in forensic, gene
CC   therapy and chromosome mapping procedures. The sequences can also be used
CC   for obtaining corresponding promoter sequences. The nucleic acids
CC   encoding the signal peptides can be used for directing extracellular
CC   secretion of a polypeptide or the insertion of a polypeptide into a
CC   membrane, or importing a polypeptide into a cell.
SQ   Sequence   89 AA;

alignment_scores:
         Quality: 452.00   Length: 89
         Ratio: 5.079   Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-030-606-111 x Y11862   ..

Align seg 1/1 to: Y11862   from: 1   to: 89

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTTCAATTGCT 172
    |||||||||||||||||||||||||||||||||||||||||||||||
  1 MetGlnCysPheSerPheIleLysThrMetMetIleLeuPheAsnLeuLe 17

173 CATCTTTCGTGTGGTGCAGCCCGTTGGCAGTGGGCATCGGGTGTCAA 222
    |||||||||||||||||||||||||||||||||||||||||||||||
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerI 34

223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCC 272
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 34 leAspGlyAlaSerPheLeuLysIlePheGlyProLeuSerSerSerAla 50

273 ATGCAGTTGTCAACGTGGGCTACTCCTCCATCGCAGCCGGCGTTGTGGT 322
    |||||||||||||||||||||||||||||||||||||||||||||||||
 51 MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyValValVa 67

323 CTTTGCCTCTGGTTCCTCGGGCTGCTATGGTGCTAAGACTGACAGCAAGT 372
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 67 lPheAlaLeuGlyPheLeuGlyCysTyrGlyAlaLysThrGluSerLysC 84

373 GTGCCCTCGTGACGTTC 389
    |||||||||||||||||
 84 ysAlaLeuValThrPhe 89

---

seq_name: A_Geneseq_36:W86331

seq_documentation_block:
ID   W86331 standard; Protein; 140 AA.
AC   W86331;
DT   01-MAR-1999 (first entry)
DE   Kidney injury associated molecule HW082 protein.
KW   Kidney injury associated molecule; kidney injury related molecule;
KW   KIM; tissue growth promotion; regeneration; renal condition;
KW   acute renal failure; acute nephritis; tumour.
OS   Rattus sp.
PN   WO9853071-A1.
PD   26-NOV-1998.
PF   22-MAY-1998;   U10547.
PR   23-MAY-1997; US-047491.
PR   23-MAY-1997; US-047490.
PA   (BIOJ) BIOGEN INC.
PI   Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
DR   WPI; 99-045312/04.
DR   N-PSDB; V80623.
PT   Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT   in injured or regenerating tissues, useful to promote tissue growth
PT   and regeneration, especially to treat renal conditions
PS   Claim 17; Page 161-162; 213pp; English.
CC   The present sequence represents a kidney injury associated molecule
CC   (KIM) protein. KIM proteins can be administered therapeutically
CC   by expressing KIM encoding polynucleotides, to promote growth and/or
CC   survival of damaged tissue (e.g. renal tissue), since the KIM proteins
CC   are upregulated in injured or regenerating (especially renal) tissues.
CC   KIM fusion proteins, conjugates, antibodies and vectors can also be used
CC   therapeutically, e.g. these or the KIM proteins may be included with an
CC   acceptable carrier in pharmaceutical compositions, useful for therapy/
CC   prophylaxis of conditions associated with disfunction/disregulation of
CC   KIM genes or proteins, especially renal diseases or impairments of renal
CC   function in humans (e.g. acute renal failure, acute nephritis). The
CC   polynucleotides can be used to produce antisense sequences which, when
CC   internalised into cells, can disrupt expression of a cellular KIM gene,
CC   also useful in therapy (e.g. to block the growth of tumours dependent on
CC   KIM for growth) or compositions. The proteins and polynucleotides are
CC   useful diagnostically e.g. to detect and quantify renal injury/disease
CC   (indicative of increased risk, or presence of, renal injury or impaired
CC   function), or abnormal responses to tissue injury (indicative of
CC   increased risk, or presence of, an autoimmune response or abnormal
CC   tissue growth arising from/affecting renal tissue). The proteins can
CC   also be used to locate KIM-producing cells (especially specific loci,
CC   e.g. tissue masses abnormally producing/expressing KIM such as tumours
CC   arising from/affecting renal tissue), by contacting cells with an
CC   imageable KIM-binding reagent and imaging reagent accumulation.
SQ   Sequence   140 AA;

alignment_scores:
         Quality: 382.00   Length: 87
         Ratio: 4.602   Gaps: 0
Percent Similarity: 95.402   Percent Identity: 83.908

alignment_block:
US-09-030-606-111 x W86331   ..

Align seg 1/1 to: W86331   from: 1   to: 140

123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTTCAATTGCT 172
    |||||||||||||::::|||||||:::||||||||||||||||||||
  1 MetGlnCysPheLysPheIleLysValMetMetIleLeuPheAsnLeuLe 17

173 CATCTTTCGTGTGGTGCAGCCCGTTGGCAGTGGGCATCGGGTGTCAA 222
    |||||||||||||||||||||||||||||||||||||||||||||:
 17 uIlePheLeuCysGlyAlaAlaLeuLeuAlaValGlyIleTrpValSerV 34

223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCC 272
    ::|||||||||||||||||||||||||||||::|||||||||||||||||
 34 alAspGlyThrSerPheLeuLysAlaPheGlySerLeuSerSerSerAla 50

```
273 ATGCAGTTTGTCAACGTGGCCTACTTCCTCATCGCAGCCGGCGGTTGTGGT 322
    |||||||||||||||||||||||||||||||||||||::|||||||::||
51  MetGlnPheValAsnValGlyTyrPheLeuIleAlaAlaGlyAlaValLe 67

323 CTTTGCCTCTTGGTTTCCGGGCGTGCTATGGTGCTAAGACTGAGAGCAAGT 372
    :||| |||||||||||||||||||||||||||| :::||||| ||||
67  uPheIleLeuGlyPheLeuGlyCysTyrGlyAlaHisSerGluAsnLysC 84

373 GTGCCCTCGTG 383
    ||:::||| ||
84  ysValLeuMet 87

seq_name: A_Geneseq_36:W61622

seq_documentation_block:
ID  W61622 standard; Protein; 238 AA.
AC  W61622;
DT  27-OCT-1998  (first entry)
DE  Clone HTBPA27 of TM4SF superfamily.
KW  Human; receptor; immune disorder; cancers; blood disorder;
KW  juvenile rheumatoid arthritis; Graves disease.
OS  Homo sapiens.
PN  WO9831799-A2.
PD  23-JUL-1998.
PF  21-JAN-1998; U00959.
PR  21-JAN-1997; US-034205.
PR  21-JAN-1997; US-034204.
PA  (HUMA-) HUMAN GENOME SCI INC.
PI  Gentz RL, Ni J, Rosen CA;
DR  WPI; 98-427559/36.
DR  N-PSDB; V48117.
PT  New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)
PT  - used to develop products for diagnosing or treating e.g. immune
PT  disorders, cancers, blood disorders or immuno-compromised disease
PT  states
PS  Claim 11; Page 44-45; 79pp; English.
CC  Clone HTBPA27 is a member of the TM4SF receptor superfamily.  The
CC  products generated using the receptor can be used for treating abnormal
CC  conditions related to both an excess of and insufficient amounts of
CC  receptor activity.  They can be used in the treatment of e.g. immune
CC  disorders, cancers, blood disorders, juvenile rheumatoid arthritis,
CC  Graves disease or immunocompromised disease states.  The products can
CC  also be used for detection and diagnosis.
SQ  Sequence 238 AA;

alignment_scores:
    Quality:   297.00      Length:    247
    Ratio:     1.916       Gaps:        9
Percent Similarity: 62.753  Percent Identity: 29.150

alignment_block:
US-09-030-606-111 x W61622

Align seg 1/1 to: W61622  from: 1  to: 238

129 TGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTTCATCTT 178
    |||::|| ::::||| :::||| ||||||||||||||||::::::
5   CysLeuGlnAlaValLysTyrLeuMetPheAlaPheAsnLeuPhePheTr 21

179 TCTGTGTGGTGCAGCCCTGCTGGCAGTGGGCATCTGGGTGTCAATCGATG 228
    :||| ||:::::::||||||||::|||||||||||||:||::::
21  pLeuGlyGlyCysGlyValLeuGlyValGlyIleTrpLeuAlaAlaThrG 38

229 GGGCATCCTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAG 278
    :::||||| :::::||||| |||||||||||::
38  lnGlySerPheAlaThrLeu.........SerSerSerPhePro 49

279 TTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGGTGTCTTTGC 328
    ::|||||| ::||||: ::||||:: |||||::
50  SerLeuSerAlaAlaAsnLeuLeuIleIleThrGlyAlaPheValMetAl 66

329 TCTTGGTTTCCTGGGCTGCTATGTGCTAAGACTGAGAGCAAGTGTGCCC 378


    |:::::|||||::|||||     ||||||
66  aIleGlyPheValGlyCysLeuGlyAlaIleAlaLysGluAsnLysCysLeuL 83

379 TCGTGACGTTCTTCATCCTCCTCCATCTTCATTGCTGCTGAGGTTGCA 428
    |:::||||||||:::||||||||||||||||:::::: ||:::::||||
83  euLeuThrPhePheLeuLeuLeuLeuLeuValPheLeuLeuGluAlaThr 99

429 GCTGCTGTGGTCGCCTTGGTGTACACCACCAATGGTGAGACTTCCTGAC 478
    ||:::: :::|||| :::::||||||||::::::
100 IleAlaIleLeuPheAlaTyrThrAspLysIleAspArgTyr...... 114

479 GTTGCTGGTAGTGCCTGCCATCAAGAAAGAT........TATGGTTCCC 519
    ::  :::||||||
115 ......AlaGlnGlnAspLeuLysLysGlyLeuHisLeuTyrGlyThrG 129

520 AGGAAGAC......TTCACTCAAGTGTGGAACACCACCATGAAAGGGCTC 563
    ||  :::||||||||:||::     ::    :::
129 lnGlyAsnValGlyLeuThrAsnAlaTrpSerIleIleGlnThrAspPhe 145

564 AAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTT 613
    ::||||||| |||||:::||||||||:::|||||||||
146 ArgCysCysGlyValSerAsnTyrThrAsp..............TrpPh 157

614 CAAAGAGAACAGTGCC......TTTCCCCATTCTGTTGCAATGACAACG 657
    ||::|    :::||||            |||||
157 eGluValTyrAsnAlaThrArgValProAspSerCysSerLeuGlu..... 172

658 TCACCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAA 707
    :::||||::|||
173 ............PheSerGluSerCysGlyLeuHisAlaProGlyThrTrp 185

708 AAAGTAGAGGTTGCTTCAATCAGCTTTGTATGACATCCGAACTAATGC 757
    |||::::::||| ::      :::     ||
186 TrpLysAlaProCysTyrGluThrValLysValTrpLeuGlnGluAsnLe 202

758 AGTCACCGTGGT......GGTGTGGCAGCTGGAATTGGGGGCCTCGAGC 801
    :::::|||||      |||:::||||       :::::
202 uLeuAlaValGlyIlePheGlyLeuCysThrAlaLeu......ValGlnI 217

802 TGGCTGCCATGATTGTGTCCATGATATCTGTACTGCAAATCTA 842
    :: :::::     |||||||     |||::|||::::::
217 leLeuGlyLeuThrPheAlaMetThrMetTyrCysGlnVal 230

seq_name: A_Geneseq_36:W05732

seq_documentation_block:
ID  W05732 standard; Protein; 267 AA.
AC  W05732;
DT  18-FEB-1997  (first entry)
DE  Human metastasis tumour suppressor gene KAI1 product.
KW  Metastasis; tumour suppressor gene; KAI1; cancer; diagnosis;
KW  gene therapy.
OS  Homo sapiens.
FH  Key              Location/Qualifiers
FT  domain           11..34
FT                   /label= Transmembrane_domain
FT  domain           55..78
FT                   /label= Transmembrane_domain
FT  domain           84..108
FT                   /label= Transmembrane_domain
FT  modified_site    129
FT                   /label= Glycosylation
FT                   /note= "putative N-glycosylation site"
FT  modified_site    157
FT                   /label= Glycosylation
FT                   /note= "putative N-glycosylation site"
FT  modified_site    198
FT                   /label= Glycosylation
FT                   /note= "putative N-glycosylation site"
PN  WO9634117-A1.
PD  31-OCT-1996.
PF  25-APR-1996; U05848.
PR  28-APR-1995; US-430225.
```

PA   (UJJO) UNIV JOHNS HOPKINS.
PA   (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI   Barrett JC, Dong J, Isaacs JT, Lamb PW.
WPI; 96-497645/49.
DR   N-PSDB; T40021.
PT   Method for detecting human metastasis suppressor gene KAI1 - useful
PT   for developing prods. for the diagnosis, prognosis and therapy of
PT   malignant cancers
PS   Example 2; Fig 3; 49pp; English.
CC   The 29.6 kDa human KAI1 protein (W05732) is the product of the
CC   human metastasis tumour suppressor gene KAI1 (T40021), and is
CC   expressed in many tissues. Recombinant KAI1 protein can be
CC   produced in transformed host cells for diagnostic and prognostic
CC   applns. Alterations in the protein sequence are indicative of the
CC   presence of malignant cancer, or of a predisposition to malignancy,
CC   in a subject. Gene therapy can be used to restore the wild-type
CC   KAI1 gene product to a subject.
SQ   Sequence   267 AA;

alignment_scores:
            Quality: 294.50   Length: 268
            Ratio: 1.763   Gaps: 11
Percent Similarity: 62.313   Percent Identity: 29.851

alignment_block:
US-09-030-606-111 x W05732

Align seg 1/1 to: W05732   from: 1   to: 267

129 TGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTT 178
       |||:|||| ||| :::::::::||||||||||:::::||
  5 CysIleLysValThrLysPheLeuPheLeuPheAsnLeuIlePhePh 21

179 TCTGTGGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATG 228
       |:::: ||||||||:::|||::: ||||:::|||:::
 21 eIleLeuGlyAlaValIleLeuGlyPheGlyValTrpIleLeuAlaAspL 38

229 GGGCATCCTTTCTGAAGATCTTCGGGCACTGTCGTCCAGTGCCATGCAG 278
       ::|||||||||||| ||||||||
 38 ysSerSerPheIleSerValLeuGlnThrSerSerSer......... 51

279 TTTGTCAACGTGGGC...TACTTCCTCATCCGAGCCGGCGGCGTTGTGGTCTT 325
       :::::::||| ||| :|||:::||||||||:::||||:|||||:
 52 ...LeuArgMetGlyAlaTyrValPheIleGlyValGlyAlaValThrMe 67

326 TGCTCTTGGTTCCTGGGCGTCGTATGGTGCTAAGACTGGAGCAAGTGTG 375
       :||||||||||||| |||||| ||||| :::|||
 67 tLeuMetGlyPheLeuGlyCysIleGlyAlaValAlaValArgCysL 84

376 CCCTCGTGACGTTCTTCTTCATCCTCCCTCATCTTCAATTCATTGCTGAGGTT 425
       :::||||||||| :|||||||||:::|||||:|||||||||
 84 euLeuGlyLeuTyrPheAlaPheLeuLeuLeuLeuIleLeuAlaGlnVal 100

426 GCAGCTGCTGGTCGTGGTCGGTGTACACCACCATGGCTGAGCACTTCCT 475
       :::||||||||||| :||||:||||||||||
 101 ThrAlaGly......AlaLeuPheTyrPheAsnMetGlyLysLeuLysGl 115

476 GACGTTG...CTGGTAGTGCCTGCCATCAAGAAAGATTAT...GGTTCCC 519
       :::::::| |||| :::|||||| :::||||| :::||||
 115 nGluMetGlyGlyIleValThrGluLeuIleArgAspTyrAsnSerSerA 132

520 AGGAAGACTTCACTCAA...GTGTGGAACACCACCATGAAAAGGGCTCAAG 566
       :::||||| ||| :::|||| :::::: :::|||
 132 rgGluAspSerLeuGlnAspAlaTrpAspTyrValGlnAlaGlnValLys 148

567 TGCTGTGGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTC... 614
       ||||||||||||||||||| ||:||| :::
 149 CysCysGlyTrpValSerPheTyrAsnTrpThrAspPAsnAlaGluLeuMe 165

615 ...AAAGAGAAACAGTGCCTTCCCCCATTCTGT...........TGCA 648
       :: :::::|||| |||||:::||||||| ||||
 165 tAsnArgProGluValThrThrTyrProCysSerCysGluValLysGlyGluG 182

649 ATGACAACGTCACCACCACACAGCCAATGAAACCTGC............ 683
       :|||||||| :::: ||
 182 luAspAsnSerLeuuSerValArgLysGlyPheCysGluAlaProGlyAsn 198

684 ...ACCAAGCAAAAGGCTCACGACCACCAAAA............GTAGAGGG 718
       :||||||||:|||| ::::||| :::
 199 ArgThrGlnSerGlyAsnHisProGluAspTrpProValTyrGlnGluGl 215

719 TTGCTTCAATCAGCTTTGTATGACATCCGAACTAATGCACCAGTCACCGTGG 768
       ||||::::::::::|| ||| |||
 215 yCysMetGlyLysValGlnAlaTrpLeuGlnGluAsnLeuGlyIleIleL 232

769 GTGGTGGCAGCTGGAATTGGGGGCCTCGAGTCGCTGCCATGATTGTG 818
       |||||:::||||||:::|||||| :::||||| :::|||:|||
 232 euGlyValGlyValGlyValAlaIleAlaIleIleGluLeuLeuGlyMetValLeu 248

819 TCCATGTATCTGTACTGCAATCTACAATAAGTCCACTTCTGCCTCTGCCA 868
       |||:::|
 249 SerIle................................CysLeuCysAr 254

869 CTAC 872
       ::::
 254 gHis 255

seq_name: A_Geneseq_36:W61624

seq_documentation_block:
ID   W61624 standard; Protein; 252 AA.
AC   W61624;
DT   27-OCT-1998 (first entry)
DE   Clone HHFEK40 of TM4SF superfamily.
KW   Human; receptor; immune disorder; cancers; blood disorder;
KW   juvenile rheumatoid arthritis; Graves disease.
OS   Homo sapiens.
PN   WO9831799-A2.
PD   23-JUL-1998.
PF   21-JAN-1998; U00959.
PR   21-JAN-1998; US-034205.
PR   21-JAN-1997; US-034204.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gentz RL, Ni J, Rosen CA;
DR   WPI; 98-427559/36.
DR   N-PSDB; V48120.
PT   New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)
PT   - used to develop products for diagnosing or treating e.g. immune
PT   disorders, cancers, blood disorders or immuno-compromised disease
PT   states
PS   Claim 11; Page 48-49; 79pp; English.
CC   Clone HHFEK40 is a member of the TM4SF receptor superfamily. The
CC   products generated using the receptor can be used for treating abnormal
CC   conditions related to both an excess of and insufficient amounts of
CC   receptor activity. They can be used in the treatment of e.g. immune
CC   disorders, cancers, blood disorders, juvenile rheumatoid arthritis,
CC   Graves disease or immunocompromised disease states. The products can
CC   also be used for detection and diagnosis.
SQ   Sequence   252 AA;

alignment_scores:
            Quality: 288.50   Length: 250
            Ratio: 1.748   Gaps: 9
Percent Similarity: 66.000   Percent Identity: 32.000

alignment_block:
US-09-030-606-111 x W61624

Align seg 1/1 to: W61624   from: 1   to: 252

138 TTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTCGTGTGG 187
       :::||| ||||||||||||||||||||:::||
 15 TyrLeuLysTyrLeuPheLeuPheAsnPhePhePheValGlyGl 31

```
188  TGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCT  237
     ||||||:::||||||||||||||||:::   :::::::  :::::::
 31  yAlaAlaValLeuAlaValGlyIleTrpLeuValGluLysSerGlyT    48

238  TTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAAC   287
     ::||:::::::::      |||::||||||::  |||:::
 48  yrLeuSerVal........LeuAlaSerThr......PheAlaAla     59

288  GTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGCTCTTGTTT   337
     ::||||::::::  ||||:::::::::||||
 60  SerAlaTyrIleLeuIlePheAlaGlyValLeuValMetValThrGlyPh  76

338  CCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGT  387
     |||||||      |||   |||   :::|||   |||
 76  eLeuGlyPheGlyAlaIleLeuTrpGluArgLysGlyCysLeuSerThrT  93

388  TCTTCTTCATCCTCCTCATCTTCATTGCTGAGGTTGCAGCTGCTGTG     437
     ::|||   ::::||||:||||:::|||||:::||:::|||:||||
 93  yrPheCysLeuLeuValIlePheLeuValGluLeuValAlaGlyVal    109

438  GTCGCCTTGGTGTACACCACAATGGCTGAGCACTTCCTGACGGTTGCTGGT 487
     ::||||  |||||   :::       |||
110  LeuAlaHisValTyrTyrGlnArgLeuSerAspGluLeuLysGlnHisLe 126

488  AGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGA......GACTTCA   531
     :    ::||||::::: :::::|||||:::::::    |||  :
126  uAsnArgThrLeuAlaGluAsnTyrGlyGlnProGluHisAlaAspHisA 143

532  CTCAAGTGTGGAACACCACCATGAAAGGCTCAAGTGCTGTGCTTCACC    581
     :::::||| :::   ::|||||||||||  :::
143  laSerVal..AspArgLeuGlnGlnAspPheLysCysCysGlySerAsn  158

582  AACTATACGGATTTTGAGGACTCACCCTACTTC...........AAAGA  619
     :::  ::||||||:::::: |||  ::::    ||
159  SerSerAlaAspTrpGlnHisSerThrTyrIleLeuLeuArgGluAlaGl 175

620  GAACAGTGCCTTCCCCATTCTGTCAATGACAACGTCACCAACACAG      669
     |::: |||  ||||||||:::::::  :
175  uGlyArgGlnValProAspSerCysLysThrValAlaArgCysG       192

670  CCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAA........AAA  710
     ||:::||||||  :::        |||
192  ly..............GlnArgAlaHisProSerAsnIleTyrLys     202

711  GTAGAG...GGTTGCTTCAATCAGCTTTGTATGACATCCGAACTAATGC   757
     ||||||   :::::::::::::||  :::    :
203  ValGluGlyGlyCysLeuThrLysLeuGlnPheLeuAlaAspHisLe    219

758  AGTCACCGTGGGTGGTGGCAGCTGGAATTGGGGGCCTCGAGCTGGCTG    807
     :::   ::||||:|||| ||||||||||| :::::::::
219  uLeuLeuMetGlyAlaValGlyIleGlyValAlaCysLeuGlnIleCysG 236

808  CCATGATTGTGTCCATGTATCTGTACTGCAATCTACAATAAGTCCACTTC  857
     ::||::::: ::|||||| ||||||
236  lyMetValLeuThrCysCysLeuHisGlnAsnArgLeuGln...ArgHisPhe 251
```

OM of: US-09-030-606-111 to: Issued_Patents_AA:*   out_format : pfs

Date: Sep 25, 1999 4:01 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09030606/runat_24091999_171616_29813/app_query.fasta.1
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
Query: US-09-030-606-111
Query length: 1289
Database: Issued_Patents_AA:*
Database sequences: 106577
Database length: 9868381
Search time (sec): 110.560000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len ! Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-855-140-1	+	285.50	529.73	7.3e-23	280
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-807-044-1	+	281.50	522.58	1.9e-22	265
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-855-140-3	+	269.00	500.45	4.0e-21	219
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-807-044-3	+	269.00	500.45	4.0e-21	219
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US91-04986-2	+	269.00	500.45	4.0e-21	219
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-855-140-4	+	263.00	488.76	1.8e-20	219
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-254-109-1	+	217.00	398.68	1.8e-15	227
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-253-751-6	+	217.00	398.68	1.8e-15	227
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-453-925-6	+	217.00	398.68	1.8e-15	227
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-408-232B-1	+	217.00	398.63	1.8e-15	228
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-102-942A-6	-	99.50	161.65	0.0146	449
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-660-963-12	+	98.00	157.51	0.0223	498
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-57-758-4	+	97.50	157.75	0.0240	449
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-909-984A-4	+	96.50	146.36	0.0463	1003
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-909-983-4	+	96.50	146.36	0.0463	1003
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-234-783-4	-	94.00	151.46	0.0563	429
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-456-907-4	+	94.00	151.46	0.0563	429
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-05523-4	-	94.00	151.46	0.0563	429
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-09261-2	+	93.00	146.83	0.0811	539
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-985-090-5	+	86.50	138.83	0.3358	362
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-127-499A-1	+	87.50	128.95	0.4366	992
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-482-847-1	+	87.50	128.95	0.4366	992
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-102-942A-2	-	84.00	134.53	0.6142	345
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US93-08386-9	-	84.50	131.80	0.6357	473
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-336-891-2	+	83.50	133.62	0.6999	343
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-13795-4	+	83.50	133.62	0.6999	343
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-313-288B-15	+	83.50	129.95	0.8128	469
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-985-090-2	+	82.50	128.61	1.02	445
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-349-025-2	+	81.00	125.40	1.50	456
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US95-15646-2	+	81.00	125.40	1.50	456
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-668-648-4	+	81.50	123.30	1.51	593
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-429-998-4	+	81.50	123.30	1.51	593
/cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US91-02321-4	+	81.50	123.30	1.51	593
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-469-537A-72	+	82.00	120.76	1.55	800
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-469-537A-78	+	82.00	120.76	1.55	800
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-568-459A-4	-	83.00	115.85	1.62	1435
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-469-537A-105	+	82.00	118.90	1.68	937
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-653-740-7	+	80.50	120.77	1.99	623
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-553-497A-26	+	78.50	127.97	2.03	242
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-026-138E-4	-	81.50	113.88	2.26	1323
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-414-926A-9	+	78.50	124.40	2.37	328
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-276-967-2	+	82.50	108.47	2.42	2476
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-046-585-5	+	82.00	109.80	2.48	2035
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-393-703-5	+	82.00	109.80	2.48	2035

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-855-140-1

seq_documentation_block:
; Sequence 1, Application US/08855140
; Patent No. 5854022
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/855,140
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0296 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MYOMNOT01
; CLONE: 779308
US-08-855-140-1

alignment_scores:
Quality: 285.50   Length: 185
Ratio: 2.399   Gaps: 7
Percent Similarity: 64.324   Percent Identity: 35.676

alignment_block:
US-09-030-606-111 x US-08-855-140-1

Align seg 1/1 to: US-08-855-140-1 from: 1 to: 280

```
129 TGCTTCAGCTTCATTAAGACCATGATGATCCTCCTTCCAATTGCTCATCTT 178
    :||||  ||||||  :::||||  |||||||:::|||||||||||||::::
  5 CysLeuCysLeuLysTyrMetMetPheLeuPheAsnLeuIlePheTr    21

179 TCTGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATG 228
    |||||||||| ||||||||||||||||||||||||||||||||:::::
 21 pLeuCysGlyCysGlyLeuLeuGlyValGlyIleTrpLeuSerValSerG  38
```

```
;; TELECOMMUNICATION INFORMATION:
;;   TELEPHONE:  415-855-0555
;;   TELEFAX:    415-845-4166
;;   TELEX:
;; INFORMATION FOR SEQ ID NO: 1:
;;   SEQUENCE CHARACTERISTICS:
;;     LENGTH: 265 amino acids
;;     TYPE:  amino acid
;;     STRANDEDNESS: single
;;     TOPOLOGY: linear
;;   IMMEDIATE SOURCE:
;;     LIBRARY: SCORNOT01
;;     CLONE: 663655
; US-08-807-044-1

alignment_scores:
        Quality:  281.50               Length:    274
        Ratio:    1.793                Gaps:       11
Percent Similarity:  57.299    Percent Identity:  27.372

alignment_block:
US-09-030-606-111 x US-08-807-044-1

Align seg 1/1  to: US-08-807-044-1  from: 1  to: 265

129 TGCTTCAGCTTCATTAAGACAATGATGATCCTCTTCAATTGCTCATCTT 178
    |||:::::   :::||||   :::||||   |||||||||||:::::
  5 CysLeuGlnAlaValLysTyrLeuMetPheAlaPheAsnLeuLeuPheTr 21

179 TCTGTGGTGGTGCAGCCCTGTGCAGTGGGCATCTGGGTGTCAATCGATG 228
    :|||    |||:::||||||||||||||||||||:::::||||
 21 pLeuGlyGlyCysGlyValLeuGlyValGlyIleTrpLeuAlaIleThrG 38

229 GGGCATCCTTTCTGAAGATCTTCGGGGCACTGTCGTCCAGTGCCATGCAG 278
    :::||||||           :::::||||
 38 lnGlySerPhe..........AlaThr***SerSerSerPhePro 49

279 TTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGC 328
    :::::::::   :::||||||   |||:::|||||  :::::
 50 SerLeuSerAlaAlaAsnLeuLeuLeuIleThrGlyAlaPheValMetAl 66

329 TCTTGGTTCCTGGCTGCTATGTGCTAAGACTGAGACAAGTGTGCCC 378
    |:|||:|||  |||||   ||||| :::||||
 66 alIleGlyPheValGlyCysLeuGlyAlaIleLysGluAsnLysCysLeuL 83

379 TCGTGACGTTCTTCATCCTCCTCCATCTTCATTGCTGAGGTTGCA 428
    |||:||||||||||   |||::  :::||||  :::||||
 83 euLeuThrPhePheLeuLeuLeuLeuValPheLeuLeuGluAlaThr 99

429 GCTGCTGGTGCGCCTGGTGTACACCAAATGGCTGAGCACTTCCTGAC 478
    :::||||||            :::||||||
100 IleAlaIleLeuPhePheAlaTyrThrAspLysIleAspLysTyr..... 114

479 GTTGCTGGTAGTGCCTGCCATCAAGAAAGAT.......TATGGTTCCC 519
    ::::|||||    ||||||:::||||
115 ......AlaGlnInAspLeuLeuLysLysGlyLeuHisLeuTyrGlyThrG 129

520 AGGAAGAC.....TTCACTCAAGTCGGAAC........ACCACATG 554
    ||   ::||     :::||||::||:::
129 lnGlyAsnValGlyLeuThrAsnAlaTrpSerIleIleGlnInThrAsp*** 145

555 AAAGGGCTC................... 563
    :::|||::
146 ArgGlyValGlyArgTrpAlaGlySerAlaGlyAlaProSerPro***Al 162

564 ..................AAGTGCTGTGGCTTCACCA 582
    :::||||||   ::|||
162 aSerAlaArgProGluLeuAlaProGlnPheArgCysCysGlyValSerA 179

583 ACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAACAGTGCC... 629
    ||||||||||        ::|||
179 snTyrThrAsp.........TrpPheGluValTyrAsnAlaThr 190
```

```
229 GGGCATCCTTTCTGAGATCTTCGGGCACTGTCGTCCAGTGCCATGCAG 278
    :::::|||          |||:::|||        |||
 38 lnGlyAsnPhe...AlaThrPheSerProSerPheProSer........ 50

279 TTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGC 328
    :::::::::   :::||||||    |||:::|||||  :::::
 51 ...LeuSerAlaAlaAsnLeuValIleAlaIleGlyThrIleValMetVa 66

329 TCTTGGTTCCTGGGCTGCTATGTGCTAAGACTGAGACAAGTGTGCCC 378
    |  :|||||||||||  ||||   :::||||
 66 lThrGlyPheLeuGlyCysLeuGlyAlaIleLysGluAsnLysCysLeuL 83

379 TCGTGACGTTCTTCTTCATCCTCCTCCATCTTCATTGCTGAGGTTGCA 428
    |  ||||||||||   |||::  :::|||||  :::||||
 83 euLeuSerPheValLeuLeuValValThrLeuLeuLeuAlaGluGluIle 99

429 GCTGCTGGTGCGCCTTGGTGTACACCACAAATGGCTGAGCACTTCCTGAC 478
    :::||||||
100 LeuProIleLeuPhePheValTyrMetAspLys................. 110

479 GTTGCTGGTAGTGCCTGCCATCAAGAAAGAT................. 509
    ||   ||||||||
111 ......ValAsnGluAsnAlaLysLysAspLeuLysGluGlyLeuLeuL 125

510 ..TATGGTTCCAGGAAGAC......TTCACTCAAGTGTGGAACACCACC 551
    |||    :|||||||        :::||||
125 euTyrHisThrGluAsnAsnValGlyLeuLysAsnAlaTrpAsnIleIle 141

552 ATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGGA 601
    :::   |||||||||||  |||||||||||
142 GlnAlaGluMetArgCysCysGlyValThrAspTyrThrAspPhe..... 156

602 CTCACCCTACTTCAAAGAGAACAGTGCCTTTCCCCATTCTGTTGCAATG 651
    |||    :::||||:::::|         |||
157 .TyrProValLeuGlyLeuAsnThrVal...ProAspArgCysCysMetG 172

652 ACAAC 656
    ::|||
172 luAsn 173
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-807-044-1

```
                              Quality:  269.00                    Length:    240
                                Ratio:    1.781                      Gaps:      5
Percent Similarity:  62.917    Percent Identity:  25.833

alignment_block:
US-09-030-606-111 x US-08-855-140-3

Align seg 1/1  to: US-09-030-606-111  x  US-08-855-140-3   from: 1  to: 219
```

```
  123 ATGCAGTGCTTCAGCTTCATTAAGACCATGATGAGATCCTCTTCAATTGCT  172
       |||:||  :::||||||||||||||||:||  |||||||  ||:||||||
    3 MetSerSerLeuLysLeuLeuLysTyrValLeuPhePheAsnLeuLeu     19

  173 CATCTTTCTGTGGTGGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA    222
       |:::::::||||||:|||:|::::||||:|||||  ||||||::||||||:
   19 uPheTrpIleCysGlyCysIleLeuGlyPheGlyIleTyrLeuLeuI     36

  223 TCGATGGGGCATCCTTCTGAAGATCTTCGGGCCCACTGTCGTCCAGTGCC    272
       ::::||     ::||||||||:::|||           |||      |||
   36 leHisAsn...AsnPheGlyValLeuPheHisAsnLeuProSer.......  49

  273 ATGCAGTTTGTCAACGTGGCTACTTCCTCATCCAGCCGGCCGTTCGTGGT   322
       ::::::||||         ::::||||||||:::::::::  :::::|||
   50 ........LeuThrLeuGlyAsnValPheValIleValGlySerIleI    63

  323 CTTTGCTCTTGTTTCTGGGCTGCTGCTATGGTGCTAAGACTGCAGAGCAAGT  372
       :::::||||||||||||||||||  |||::  ||::  ||::::::||||
   63 eMetValValAlaPheLeuGlyCysMetGlySerIleLysGluAsnLysC   80

  373 GTGCCCTCGTGACGTTCTTCTCCATCCTCCTCCATCCTCCATTGCTGAG    422
       |||  ::::::|||||||||::::|||::|::|::||||||||:||||||
   80 ysLeuLeuMetSerPhePheIleLeuLeuIleLeuLeuLeuAlaGlu     96

  423 GTTGCAGCTGCTGTGGTCGCGTTGGTGTACACCACAATGGCTGAGCACTT   472
       |||::  :::|||||::|::|||:::|||||||||||||||||||||||
   97 ValThrLeuAlaIleLeuLeuPheValTyrGluGlnLysLeuAsnGluTy   113

  473 CCTGACGTTGCTGGTAGTGCCGCTGCCATCAAGAAAGATTATGGTTCCCAGG  522
       :::::||| ::::::::||:  :::||||||||||||  |||:::||||
  113 rValAlaLysGlyLeuThrAspSerIleHisArg...TyrHisSerAspA   129

  523 AAGACTTCACTCAAGTGTGAACACCACCATGAAAGGGCTCAAGTGCTGT    572
       :::: ::::::::||||::||||:||:|||::::||:::||||||||
  129 snSerThrLysAlaAlaTrpAspSerIleGlnSerPheLeuGlnCysCys  145

  573 GGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTTCAAAGAGAA   622
       :::||::::::||:|||| |||||||||||||||::::||||||||||
  146 GlyIleAsnGlyThrSerAspTrp......................     153

  623 CAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCACCAACACAGCCA   672
       ::::|| ||||   |||||  |||||||||  :::
  154 .ThrSerGlyProProAlaSerCys................     161

  673 ATGAAACCTGCACCAAGCAAGAGGCTCACGACCAAAAAGTAGAGGGTTGC  722
                          |||::|||||||||||||||||
  162 ..............ProSerAspArgLysValGluGlyCys         170

  723 TTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCACCGTGGGTGG   772
       :::: :::::::::::::|||:::::|||| :::  ::::||||||||
  171 TyrAlaLysAlaArgLeuTrpPheHisSerAsnPheLeuTyrIleGlyIl  187

  773 TGTGCCAGCTGGAATTGGGGGCCCTGCAGCTGGCTCGCCATGATTGTCTCCA  822
       ::::::  :::::|||:||:::||  ::||||||||||  :::
  187 eIleThrIleCysValCysValIleGluValLeuGlyMetSerPheAlaL  204

  823 TGTATCTGTACTGCAATCTA  842
       ||| |||  ||:::::||
  204 euThrLeuAsnCysGlnIle  210
```

```
  630 ...TTTCCCCCATTCTGTTGCAATGACAACGTCACCAACACAGCCAATGA  676
         |||   ||||||  :: ::          :::||
  191 ArgValProAspSerCysLeuGlu..........PheSerGl         202

  677 AACCTGCACCAAGCAAAGGCTCACGACCAAAAAGTAGAGGGTTGCTTCA  726
       |:::::||  :::::::||||||||||||||||||||||  |||:::||
  202 uSerCysGlyLeuHisAlaProGlyThrTrpTrpLysAlaProCysTyrG  219

  727 ATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGT......  770
       :: ::  :::::::::|||  |||          :::||||||
  219 luThrValLysValTrpLeuGlnGluAsnLeuLeuAlaValGlyIlePhe  235

  771 GGTGTGGCAGCTGGAATTGGGGGCCCTGCAGCTGGCTGCCATGATTGTGTC  820
       |||  :::::||||||||  ::::::::  ::::::|  :::::
  236 GlyLeuCysThrAlaLeu.....ValGlnIleLeuGlyLeuThrPheAl  250

  821 CATGTATCTGTACTGCAATCTA  842
       :::|  |||  |||:::|||
  250 aMetThrMetTyrCysGlnVal  257
```

```
US-08-855-140-3

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-855-140-3

seq_documentation_block:
; Sequence 3, Application US/08855140
; Patent No. 5854022
GENERAL INFORMATION:
  APPLICANT:  Hillman, Jennifer L.
  APPLICANT:  Bandman, Olga
  APPLICANT:  Goli, Surya K.
  APPLICANT:  Guegler, Karl J.
  TITLE OF INVENTION:  NEW TRANSMEMBRANE PROTEIN
  NUMBER OF SEQUENCES:  4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE:  Incyte Pharmaceuticals, Inc.
    STREET:  3174 Porter Drive
    CITY:  Palo Alto
    STATE:  CA
    COUNTRY:  USA
    ZIP:  94304
  COMPUTER READABLE FORM:
    MEDIUM TYPE:  Diskette
    COMPUTER:  IBM Compatible
    OPERATING SYSTEM:  DOS
    SOFTWARE:  FastSEQ for Windows Version 2.0
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER:  US/08/855,140
    FILING DATE:  Herewith
    CLASSIFICATION:  435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:
    FILING DATE:
  ATTORNEY/AGENT INFORMATION:
    NAME:  Billings, Lucy J.
    REGISTRATION NUMBER:  36,749
    REFERENCE/DOCKET NUMBER:  PF-0296 US
  TELECOMMUNICATION INFORMATION:
    TELEPHONE:  415-855-0555
    TELEFAX:  415-845-4166
    TELEX:
  INFORMATION FOR SEQ ID NO:  3:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  219 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    IMMEDIATE SOURCE:
      LIBRARY:  GenBank
      CLONE:  180141
```

```
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-807-044-3

seq_documentation_block:
; Sequence 3, Application US/08807044
```

```
alignment_scores:
```

```
; Patent No. 5863735
; GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Goli, Surya K.
;    TITLE OF INVENTION:  NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94304
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/807,044
;       FILING DATE:  Herewith
;       CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  PF-0224 US
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;       TELEX:
; INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  219 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;       LIBRARY:  GenBank
;       CLONE:  180141
US-08-807-044-3
```

```
alignment_scores:
        Quality:   269.00         Length:     240
          Ratio:    1.781           Gaps:       5
Percent Similarity:   62.917     Percent Identity:  25.833

alignment_block:
US-09-030-606-111 x US-08-807-044-3

Align seg 1/1  to: US-08-807-044-3  from:  1  to: 219

123 ATGGAGTGCTTCAGCTTCATTAGACACCATGATGATCCTCCTTCAATTTGCT 172
    |||::: :::::::::::||| :::::::::||||||||||||||||||
  3 MetSerSerLeuLysLeuLeuLysTyrValLeuPhePheAsnLeuLe  19

173 CATCTTTCTGTGTGGTGCGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAA 222
    ::::::::|||||||::::::::||| ::|||||:::::: |
 19 uPheTrpIleCysIleCysIleCysValLeuGlyPheGlyIleTyrLeuLeuI  36

223 TCGATGGGGCATCCTTTCTGAAGATCTTCGGGCCACTCGTCCGTCCAGTGCC 272
    || :: :|||| ::||| :|||| ||| ||| |||
 36 leHisAsn...AsnPheGlyValLeuPheHisAsnLeuProSer......  49

273 ATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGT 322
    :::::::|||| ::::: ||| :::|| :::::
 50 ......LeuThrLeuGlyAsnValPheValIleValGlySerIleIl  63

323 CTTTGCTCTTGGTTTCCTGGCGTGCTATGGTGCTAAGACTGAGAGCAAGT 372
    :::::::::::||||||||| |||::::|||||
```

```
 63 eMetValAlaPheLeuGlyCysMetGlySerIleLysGluAsnLysC  80

373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCCTCATCCTCATTCGTGAG 422
    || |||:::::|||||:::: |||||||||::|:::::|||||
 80 ysLeuLeuMetSerPhePheIleLeuLeuLeuIleLeuLeuLeuAlaGlu  96

423 GTTGCAGCTGCTGTGGTCGCCTGGTGTACACCACAATGGCTGAGCACTT 472
    |||::: |||::::: ||||:::::::||||
 97 ValThrLeuAlaIleLeuLeuPheValTyrGluGlnLysLeuAsnGluTy 113

473 CCTGACGTTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGG 522
    :::::::|||||| ::||| :::|| ||| ||||||
113 rValAlaLysGlyLeuThrAspSerIleHisArg...TyrHisSerAspA 129

523 AAGACTTCACTCAAGTGTGGAACACCACCATGAAAGGGCTCAAGTGCTGT 572
    :::: :::||||::::: ::::: |||::::||||
129 snSerThrLysAlaAlaTrpAspSerIleGlnSerPheLeuGlnCysCys 145

573 GGCTTCACCAACTATACGGATTTGAGGACTCACCCTACTTCAAAGAGAA 622
    |||::::::||| :::::||| :: :||||
146 GlyIleAsnGlyThrSerAspTrp............... 153

623 CAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCACCAACACAGCCA 672
    ::::: |||||| |||
154 .ThrSerGlyProProAlaSerCys.............. 161

673 ATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTGC 722
    ||| :::||||||||||||||
162 ...............ProSerAspArgLysValGluGlyCys 170

723 TTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGCTGG 772
    ::: :: |||||:::::::||| :: :|||
171 TyrAlaLysAlaArgLeuTrpPheHisSerAsnPheLeuTyrIleGlyIl 187

773 TGTGGCAGCTGGAATTGGGGGCCTCGAGCTGCTGCCATGATTGTGTCCA 822
    ::::: ::: :::||||:::: :::|||
187 eIleThrIleCysValCysValIleGluValLeuGlyMetSerPheAlaL 204

823 TGTATCTGTACTGCAATCTA 842
    :: ||| |||:::::::|
204 euThrLeuAsnCysGlnIle 210
```

```
seq_name: /cgn2_6/ptodata/2/iaa/PCTUS9_COMB.pep:PCT-US91-04986-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9104986
; GENERAL INFORMATION:
;    APPLICANT:  Seed, Brian
;    APPLICANT:  Allen, Janet
;    APPLICANT:  Aruffo, Alejandro
;    APPLICANT:  Camerini, David
;    APPLICANT:  Lauffer, Leander
;    APPLICANT:  Oquendo, Carmen
;    APPLICANT:  Simmons, David L.
;    APPLICANT:  Stamenkovic, Ivan
;    APPLICANT:  Stengelin, Siegfried
;    APPLICANT:  Amiot, Martine
;    TITLE OF INVENTION:  Rapid Immunoselection Cloning Method
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Greenlee & Associates
;       STREET:  5370 Manhattan Circle, Suite 201
;       CITY:  Boulder
;       STATE:  Colorado
;       COUNTRY:  USA
;       ZIP:  80303
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US91/04986
```

```
;    FILING DATE: 19910715
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US/07/553,759
;    FILING DATE: 13-JUL-1990
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 07/498,809
;    FILING DATE: 23-MAR-1990
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 07/379,076
;    FILING DATE: 13-JUL-1989
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER: US 07/160,416
;    FILING DATE: 25-FEB-1988
;  ATTORNEY/AGENT INFORMATION:
;    NAME: Wall, Margaret M.
;    REGISTRATION NUMBER: 33,462
;    REFERENCE/DOCKET NUMBER: 11-88C
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE: 303-499-8080
;    TELEFAX: 303-499-8089
;  INFORMATION FOR SEQ ID NO: 2:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 219 amino acids
;    TYPE: AMINO ACID
;    TOPOLOGY: linear
;    MOLECULE TYPE: protein
PCT-US91-04986-2

alignment_scores:
       Quality:  269.00          Length:      240
       Ratio:      1.781          Gaps:          5
Percent Similarity:  62.917   Percent Identity: 25.833

alignment_block:
US-09-030-606-111 x PCT-US91-04986-2   ..

Algn seg 1/1   to: PCT-US91-04986-2   from: 1   to: 219

123 ATGCAGGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCT 172
     |||::    :::::::::|||   :::::::::|||||||||||
  3 MetSerSerLeuLysLeuHisLysTyrValLeuPhePheAsnLeuLe  19

173 CATCTTTCTGTGTGGTGCAGCCCTGTTGGCAGTGGCATCTGGGTGTCAA 222
     ::::::::|||::||||::||||||||::::::  ||||||::::::
 19 uPheTrpIleCysGlyCysCysIleLeuGlyPheGlyIleTyrLeuLeuI  36

223 TCGATGGGGCATCCTTCTGAAGATCTTCGGGCCACTGTCGTCCAGTGCC 272
     |:|||::  :|||  :::||  ||| |||  ||||||:::::
 36 leHisAsn..AsnPheGlyValLeuPheHisAsnLeuProSer......  49

273 ATGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGCCGGTTGTGGT 322
     ::::::::|||    :::::  ||||  :::|| |||||::::||
 50 .......LeuThrLeuGlyAsnValPheValIleValGlySerIleII  63

323 CTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGT 372
     :::::::||||||||| |||::::||   |||::|||
 63 eMetValAlaPheLeuGlyCysMetGlySerIleLysGluAsnLysC  80

373 GTGCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTGAG 422
     ||  |||:::|||||:::|||||||||||::::::||||||
 80 ysLeuLeuMetSerPhePheIleLeuLeuIleIleLeuLeuAlaGlu  96

423 GTTGCAGCTGCTGTGGTGTCGCCTTGGTGTACCACAAATGCTGAGCACTT 472
     |||::::||| ::|||||               :::::::|||
 97 ValThrLeuAlaIleLeuLeuPheValTyrGluGlnLysLeuAsnGluTy 113

473 CCTGACGTTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGG 522
     ::::::||  :::||| ||||| ||| |||   :::::
113 rValAlaLysAlaGlyLeuThrAspSerIleHisArg...TyrHisSerAspA 129

523 AAGACTTCACTCAAGTGTGGAACACCACCATGAAAGGGCTCAAGTGCTGT 572
```

```
        ::::::::     :::|||||:::::    :::::::||||   |||::||||||
129 snSerThrLysAlaAlaTrpAspSerIleGlnSerPheLeuGlnCysCys 145

573 GGCTTCACCAACTATACGGATTTGAGGACTCACCCTACTTCAAAGAGAA 622
     |||::||     :::::::  ::::::::
146 GlyIleAsnGlyThrSerAspTrp...................... 153

623 CAGTGCCTTCCCCATTCTGTTGCAATGACAAGTCACCAACACAGCCA 672
     ::::::  ||||||    |||
154 .ThrSerGlyProProAlaSerCys..................... 161

673 ATGAAACCTGCCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTGC 722
     ||:::||||||||||||||||
162 ............................ProSerAspArgLysValGluGlyCys 170

723 TTCAATCAGCTTTGTATGACATCCGAACTAATGCACTCACCGTCACCGTGG 772
     :::    ||||   :::::::||||
171 TyrAlaLysAlaArgLeuTrpPheHisSerAsnPheLeuTyrIleGlyIl 187

773 TGTGGCAGCTGGAATTGGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCA 822
     :::::::  ::: :::||||::  :::|||
187 eIleThrIleCysValCysValIleGluValLeuGlyMetSerPheAlaL 204

823 TGTATCTGTACTGCAATCTA 842
     |||  ||:::::::
204 euThrLeuAsnCysGlnIle 210

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-855-140-4

seq_documentation_block:
;  Sequence 4, Application US/08855140
;  Patent No. 5854022
;  GENERAL INFORMATION:
;    APPLICANT: Hillman, Jennifer L.
;    APPLICANT: Bandman, Olga
;    APPLICANT: Goli, Surya K.
;    APPLICANT: Guegler, Karl J.
;    TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
;    NUMBER OF SEQUENCES: 4
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Incyte Pharmaceuticals, Inc.
;    STREET: 3174 Porter Drive
;    CITY: Palo Alto
;    STATE: CA
;    COUNTRY: USA
;    ZIP: 94304
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE: Diskette
;    COMPUTER: IBM Compatible
;    OPERATING SYSTEM: DOS
;    SOFTWARE: FastSEQ for Windows Version 2.0
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/855,140
;    FILING DATE: Herewith
;    CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:
;    FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;    NAME: Billings, Lucy J.
;    REGISTRATION NUMBER: 36,749
;    REFERENCE/DOCKET NUMBER: PF-0296 US
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE: 415-855-0555
;    TELEFAX: 415-845-4166
;    TELEX:
;  INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;    LENGTH: 219 amino acids
;    TYPE: amino acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;  IMMEDIATE SOURCE:
```

APPLICANT: Freeman, Gordon J.
TITLE OF INVENTION: METHODS FOR SELECTIVELY STIMULATING
TITLE OF INVENTION: PROLIFERATION OF T-CELLS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
  ADDRESSEE: LAHIVE & COCKFIELD
  STREET: 60 STATE STREET, SUITE 510
  CITY: BOSTON
  STATE: MA
  COUNTRY: USA
  ZIP: 02109
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/253,751
  FILING DATE: 3 JUNE 1994
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/073,223
  FILING DATE: 4 JUNE 1993
  APPLICATION NUMBER: US 08/200,947
  FILING DATE: 23 FEB 1994
  APPLICATION NUMBER: US 07/864,805
  FILING DATE: 7 APR 1992
  APPLICATION NUMBER: US 08/247,505
  FILING DATE: 23 MAY 1994
  APPLICATION NUMBER: US 07/864,866
  FILING DATE: 7 APR 1992
  APPLICATION NUMBER: 08/218,155
  FILING DATE: 25 MAR 1994
  APPLICATION NUMBER: US 07/864,807
  FILING DATE: 7 APR 1992
  APPLICATION NUMBER: US 07/902,467
  FILING DATE: 16 JUNE 1992
  APPLICATION NUMBER: US 07/275,433
  FILING DATE: 23 NOV 1988
ATTORNEY/AGENT INFORMATION:
  NAME: MANDRAGOURAS, AMY E.
  REGISTRATION NUMBER: 36,207
  REFERENCE/DOCKET NUMBER: RPI-002CPB
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617) 227-7400
  TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 227 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-253-751-6

alignment_scores:
          Quality: 217.00   Length: 253
          Ratio: 1.507   Gaps: 8
Percent Similarity: 56.917   Percent Identity: 27.668

alignment_block:
US-09-030-606-111 x US-08-253-751-6

Align seg 1/1 to: US-08-253-751-6   from: 1 to: 227

105 CCCTGAACGAGGAGCCACCAGCAGTGCTTCAGCTTCATTAAGACCATGAT 154
    ||| ||||||| ||::::|| ::::|| ||||| :::::|
  1 ProValLysGlyGlyThr......LysCys.........IleLysTyrLeuLe 13

155 GATCCTCCTCAATTGCTCATCTTTCTGTGTGGTGCAGCCCTGTTGGCAG 204
    ::::: |||||:::||| ::: || |||::|||||||
 13 uPheGlyPheAsnPheIlePheTrpLeuAlaGlyIleAlaValLeuAlaI 30

205 TGGGCATCTGGGTGTCAATGGGGCATCCTTTCTGAAGATCTTCGGG 254
    TGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTGAAGATCTTCGGG 254

seq_documentation_block:
; Sequence 6, Application US/08253751
; Patent No. 5858358
; GENERAL INFORMATION:
;   APPLICANT: June, Carl H.
;   APPLICANT: Thompson, Craig B.
;   APPLICANT: Nabel, Gary J.
;   APPLICANT: Gray, Gary S.
;   APPLICANT: Rennert, Paul D.

```
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/453,925
    FILING DATE: 30 MAY 1995
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/253,751
    FILING DATE: 3 JUNE 1994
    APPLICATION NUMBER: US 08/073,223
    FILING DATE: 4 JUNE 1993
    APPLICATION NUMBER: US 08/200,947
    FILING DATE: 23 FEB 1994
    APPLICATION NUMBER: US 07/864,805
    FILING DATE: 7 APR 1992
    APPLICATION NUMBER: US 08/247,505
    FILING DATE: 23 MAY 1994
    APPLICATION NUMBER: US 07/864,866
    FILING DATE: 7 APR 1992
    APPLICATION NUMBER: 08/218,155
    FILING DATE: 25 MAR 1994
    APPLICATION NUMBER: US 07/864,807
    FILING DATE: 7 APR 1992
    APPLICATION NUMBER: US 07/902,467
    FILING DATE: 16 JUNE 1992
    APPLICATION NUMBER: US 07/275,433
    FILING DATE: 23 NOV 1988
ATTORNEY/AGENT INFORMATION:
    NAME: MANDRAGOURAS, AMY E.
    REGISTRATION NUMBER: 36,207
    REFERENCE/DOCKET NUMBER: RPI-002CPB
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (617) 227-7400
    TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
    LENGTH: 227 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
    MOLECULE TYPE: protein
US-08-453-925-6
```

```
alignment_scores:
          Quality:  217.00              Length:    253
          Ratio:      1.507             Gaps:        8
Percent Similarity: 56.917     Percent Identity:  27.668

alignment_block:
US-09-030-606-111 x US-08-453-925-6   ..

Align seg 1/1  to: US-08-453-925-6   from: 1  to: 227
```

```
105 CCCTGAACAGGAGCCACCATGCAGTCAGTGCTTCAGCTTCATTAAGACCATGAT 154
    |||  |||:|||:|||              :::|||   ||||||  :::|||
  1 ProValLysGlyGlyThr...LysCys.........IleLysTyrLeuLe 13

155 GATCCTCTTCAATTGCTCATCTTCTGTGGTGCAGCCCTGTTGGCAG 204
    ::::  |||||||:::||||||||:||| :||  ||:::||||||
 13 uPheGlyPheAsnPheIlePheTrpLeuAlaGlyIleAlaValLeuAlaI 30

205 TGGGCATCTGGGTGTCAATGATGGGCATCCTTTCTGAAGATCTTCGGG 254
    ::|||:||:|||: ::||||::  ||||||  |||||||::::||||
 30 leGlyLeuTrpLeuArgPheAspSerGlnThrLysSerIlePheGluGln 46

255 CCACTGCTGCCAGTGCCATGCAGTTGTCAACGTGGCTACTTCCTCAT 304
    :::||::||::||          :::|||   |||||||::::||||
 47 GluThrAsnAsnAsnSerPheTyrThrGlyValTyrIleLeuIl 63

305 CGCAGCCGGGTGTGGTCTTGTCTTCTTGGCTGGTTCCTGGGCTGCATGGTG 354
    |::||||||::||||||||:::|||||||||||||||    ||||
```

```
 30 leGlyLeuTrpLeuArgPheAspSerGlnThrLysSerIlePheGluGln 46
255 CCACTGCTCCAGTGCCATGCAGTTGTCAAGCTGGCTACTTCCTCCTC 304
 47 GluThrAsnAsnAsnSerSerAsnSerPheTyrThrGlyValTyrIleLeuIl 63
305 CGCAGCCGCGGTGTCTGGGTCTTTGCTCTCTTCCTGGGCTGCTGCATGGTG 354
 63 eGlyAlaGlyAlaLeuMetLeuValGlyPheLeuGlyLeuGlyCysGlyA 80
355 CTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCATCCTCCTC 404
 80 laValGlnGluSerGlnCysMetLeuGlyLeuPheGlyLeuPheLeuLeu 96
405 CTCATCTTCATTGCTGAGGTTGCAGCTGCTGTGTGGTCCGCCTTGGTGTACAC 454
 97 ValIlePheAlaIleGluIleAlaAlaAlaIleTrpGly...TyrSe 111
455 CACAATGGCTGAGCACTTCCTGACGTTGCTGGTAGTGCCTGCCATCAAGA 504
111 rHisLysAspGlu.......ValIleLysGluValGlnGluPheTyrL 125
505 AAGAT...TATGGTTCCCAGGAAGACTTCACTCAAGTGTCGGAACACCACC 551
125 ysAspThrTyrAsnLysLeuLysThrLysAspGluProGlnArgGluThr 141
552 ATGAAA..........GGGCTCAAGTGCTGTGCTCCACCAACTATAC 589
142 LeuLysLysAlaIleHisTyrAlaLeuAsnCysCysGlyLeuAlaGly..... 156
590 GGATTTGAGGACTCACCCTACTTCAAAGAGAACAAGTGCCTTTCCCCCAT 639
156 ........................................... 156
640 TCTGTTGCAATGACACGTCACCAACACAGCCAATGAAACCTGCACCAAG 689
157 .........GlyValGluGlnPheIleSerAspIleCysProLys 168
690 CAAAAGGCTCACGACCAA..AAAGTAGAGGGTTGCTTCAATCAGCTTTT 736
169 LysAspValLeuGluGluThrPheThrValLysSerCysProAspAlaIleLy 185
737 GTATGACATCCGAACTAATGCAGTCACCGTGGGTGTGTGGCAGCTGGAA 786
185 sGluValPheAspAspAsnLysPheHisIleIleGlyAlaValGlyIleGlyIl 202
787 TTGGGGCCCTCGAGCTGGCTGCCATGAGTTGTCCCATGATCGTGACTGC 836
202 leAlaValValMetIleGlyMetIleLeuPheSerMetIleLeuLeuCysCys 218
837 AATCTACAA 845
219 AlaIleArg 221
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-453-925-6

63 eGlyAlaGlyLeuMetLeuValGlyPheLeuGlyCysGlyA 80

355 CTAAGACTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTC 404
    | ||||||::||| ||| ||||| ::||||||
80 laValGlnGluSerGlnCysMetLeuGlyLeuPheGlyPheLeuLeu 96

405 CTCATCTTCATTGCTGAGGTTGCAGCTGCTGCGTCGCCTTGGTGTACAC 454
    ::||||| ||| ||:||||||::| ::| ||::| TyrSe 111
97 ValIlePheAlaIleGluIleLeuAlaAlaAlaIleTrpGly......TyrSe 111

455 CACAATGGCTGAGCACTTCCTGACGTTGCGACGTGCCTGCCATCAAGA 504
    ||| :::::: ||| ::|
111 rHisLysAspGlu..........ValIleLysGlnGluPheTyrL 125

505 AAGAT...TATGGTTCCCAGGAGACTTCACTCAAGTGTGGAACACCACC 551
    ||||| :::: ::
125 ysAspThrTyrAsnLysLysThrLysAspGluProGlnArgGluThr 141

552 ATGAAA.......GGGCTCAAGTGCTGTGGCTTCCACCAACTATAC 589
    ::||| :::||| ::|||||||||||||
142 LeuLysAlaIleHisTyrAlaLeuAsnCysCysGlyLeuAlaGly...... 156

590 GGATTTTGAGGACTCACCCTACTTCAAAGAAGAACAGTGCCTTTCCCCAT 639
156 156

640 TCTGTTGCAATGACAAGTCACCAACACAGCCAATGAAACCTGCACCAAG 689
    ::||| ::: ||||| || |||
157 ..........GGTGTCGGTGCAGTTGCTTGTTGTGTGTGCAGCTGGAA 786

690 CAAAAGGCTCACGACCAA...AAAGTAGAGGGTTGCTTCAATCAGCTTTT 736
    ::: ||| |||||::|| ::||| ||||
169 LysAspValLeuGluThrPheThrValLysSerCysProAspAlaIleLy 185

737 GTATGACATCGAACTAATGCAGTCACCGTGGGTGTGTGCAGCTGGAA 786
    ::: :::::: ::||||||::||: |||
185 sGluValPheAspAsnLysPheHisIleIleGlyAlaValGlyIleGlyI 202

787 TTGGGGCCTCGAGCTGCGTGCCATGATGTGTCCATGTATCTGTACTGC 836
    ::|||| ::||||| |||| ||
202 leAlaValValMetIlePheGlyMetIlePheSerMetIleLeuCysCys 218

837 AATCTACAA 845
    ::||::
219 AlaIleArg 221

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-408-222B-1

seq_documentation_block:
; Sequence 1, Application US/08408222B
; Patent No. 5776727
; GENERAL INFORMATION:
; APPLICANT: Ikeyama, Shuichi
; APPLICANT: Koyama, Masaru
; APPLICANT: Miyake, Masauyki
; APPLICANT: Senoo, Masaharu
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Dike, Bronstein, Roberts & Cushman
;   STREET: 130 Water Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/408,222B
;   FILING DATE: 22-MAR-1995

; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/254,493
;   FILING DATE: 06-JUN-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP-079996-1991
;   FILING DATE: 12-APR-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP-085396-1991
;   FILING DATE: 14-APR-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP-022321-1992
;   FILING DATE: 07-FEB-1994
; ATTORNEY/AGENT INFORMATION:
;   NAME: Resnick, David S.
;   REGISTRATION NUMBER: 34,235
;   REFERENCE/DOCKET NUMBER: 41777-DIV
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617-523-3400
;   TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 228 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-408-222B-1

alignment_scores:
     Quality: 217.00   Length: 253
     Ratio: 1.507   Gaps: 8
Percent Similarity: 56.917   Percent Identity: 27.668

alignment_block:
US-09-030-606-111 x US-08-408-222B-1

Align seg 1/1 to: US-08-408-222B-1 from: 1 to: 228

105 CCCTGAACAGGAGCCACCATGCAGTGCTTCAGCTTCATTAAGACCATGAT 154
    ||| |||:::||| :::||| |||||| :||||| :TyrLeuLe 14
2   ProValLysGlyThr...LysCys..........IleLysTyrLeuLe 14

155 GATCCTCTCAATTGCTCATCTTCTGTGGTGGCAGCCCGTGGCAG 204
    ::||||::::|| |||::|::|:|::|::||||||||||||
14 uPheGlyPheAsnPheIlePheTrpLeuAlaGlyIleAlaValLeuAlaI 31

205 TGGGCATCTGGGTGTCAATCGATGGGCATCCTTTCTGAAGATCTTCGGG 254
    ::||||:||:: ::||||| ::: ||||||||||
31 leGlyLeuTrpLeuArgPheAspSerGlnThrLysSerIlePheGluGln 47

255 CCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGCTACTTCCTCAT 304
    :::|||| ::|| ::| ||:||||||
48 GluThrAsnAsnAsnSerPheTyrThrGlyValTyrIleLeuIl 64

305 CGCAGCCGGCGTTGTCGGTTCTTGTCCTTGGTTCCTGCTATGGTG 354
    |:::|||||:||| ::|||||||||||| ::||||||:|||:|
64 eGlyAlaGlyLeuAlaLeuMetLeuValGlyPheLeuGlyCysGlyA 81

355 CTAAGACTGAGAGCAAGTGCCCTCGTGACGTTCTTCTCATCCTCCTC 404
    || |||||||::||| ||| ||||| ||||||
81 laValGlnGluSerGlnCysMetLeuGlyLeuPheGlyPheLeuLeu 97

405 CTCATCTTCATTGCTGAGGTTGCAGCTGCTGCGTCGCCTTGGTGTACAC 454
    ::||||| ||| ||:||||||::||||||:::   :|| ||TyrSe 112
98 ValIlePheAlaIleGluIleLeuAlaAlaAlaIleTrpGly.....TyrSe 112

455 CACAATGGCTGAGCACTTCCTGACGTTGCGGTAGTGCCTGCCATCAAGA 504

REFERENCE/DOCKET NUMBER: MIT-5194A2
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617-861-6240
  TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
  LENGTH: 449 amino acids
  TYPE: amino acid
  STRANDEDNESS: not relevant
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-102-942A-6

alignment_scores:
  Quality: 99.50   Length: 452
  Ratio: 0.572   Gaps: 25
Percent Similarity: 38.496   Percent Identity: 19.469

alignment_block:
US-09-030-606-111/rev x US-08-102-942A-6

Align seg 1/1 to: US-08-102-942A-6 from: 1 to: 449

1086 GGCTACCTTAGAGGCTCGGAATGCCCCCACCCATCCACCACCAATGG 1037
     ||| ||| ||| || |::: ||| ||||||| :|||||
  49 GlySerLeuGlyGlyProAlaProProProAlaProProProProPro 65

1036 AAGGAAAGTCAGGCATCGCTAAAAGGAG......TGGTCCTATCTA 996
        :::|||||||||| :: ||| ::|
  65 oProProProHisSerPheIleLysGlnGluProSerTrpGlyGlyAlaG 82

995 GCCCC.. 991
     ::|||
  82 luProHisGluGlnLysLeuSerAlaPheThrValHisPheSerGly 98

990 ............AAGTCTGGAGCAGAAAAGGGCAGGTCCATTCTGGCCCAA 953
99 GlnPheThrGlyThrAlaGlyAlaCysArgTyrGlyProPhe......... 112

952 GTGACATTGTTAGATCCTGTCCCCTCCCCAATCACTGCTGCTGCCAGG 903
    ::||||||||| :: || ||| :||
113 ....GlyProProProSerGlnAlaSerSerG 123

902 GTGCCTCTTCACAGTTCCA.................TGTGGCAGC 874
    ||:: :||:||||| ||| ::::
123 lyGlnAlaArgMetPheProAsnAlaProTyrLeuProSerCysLeuGlu 139

873 AGTAGT.......GGCAGAGGCAGAAGTGGACTTATTCTAGATTG 836
    |||:::   :::||| ||| :: || |||||
140 SerGlnProAlaIleArgArgAsnGlnGlyTyrSerThrValThrPheAsp.G 156

835 CAGTACAGATACATGGACAC.......AATCATGGCAGCCAGCTCGAG 795
    ::::: ::||||| ::|||:::||||::
156 lyThrProSerTyrGlyHisThrProSerHisHisAlaAlaGlnPhe... 171

794 GCCCCCAATTCCAGC.................................. 780
    |||||| |||
172 ...ProAsnHisSerPheLysHisGluAspProMetGlyGlnGlnGlySe 187

779 ....TGCCCACA 773
                                      ||||||
187 rLeuGlyGluGlnGlnTyrSerValProProProValTyrGlyCysHisT 204

772 CCACCCACGGTGACTGCATTAGTTCGGATGTCATACAAAAGCTGATTGAA 723
    || ::::||| ||| |||:::
204 hrProThrAspSerCysThrGlySer.........GlnAlaLeuLeuLeu 217

722 GCAACCCCTTAC..........TTTTTGGTCGTGAGCCTTTTG... 690
    ||| || ||||||
218 ArgThrProTyrSerSerAspAsnLeuTyrGlnMetThrSerGlnLeuGl 234

689 ...................................CTTGGTGCAGGTTTCATTGGCTGTG 665

---

112 rHisLysAspGlu.........ValIleLysGluValGlnGluPheTyrL 126
    :::::: :|| ||| |

505 AAGAT...TATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACC 551
    ||||| |||:::::|| :: |||

126 ysAspThrTyrAsnLysLeuLysThrLysAspGluProGlnArgGluThr 142

552 ATGAAA..........GGGCTCAAGTGCTGTGGCTTCACCAACTATAC 589
    ::||| ||||||::||||||::
143 LeuLysAlaIleHisTyrAlaLeuAsnCysCysGlyLeuAlaGly..... 157

590 GGATTTTGAGGACTCACCCTACTTCAAAGAGAGAACAGTGCCTTTCCCCAT 639

157 157

640 TCTGTTGCAATGACAACGTCACCAACACAGCCAATGAAAACCTGCACCAAG 689
    ::||| :::|| ||| ||||
158 ....GlyValGluGlnPheIleSerAspIleCysProLys 169

690 CAAAAGGCTCACGACCAA...AAAGTAGAGGTTGCTTCAATCAGCTTTT 736
    ::: ||::||::|||||:: ||||||||| ||| |||
170 LysAspValLeuGluThrPheThrValLysSerCysProAspAlaIleLy 186

737 GTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGTGTGGCAGCTGGAA 786
    ::: :::::   :|||||:::|||::: ||||
186 sGluValPheAspAsnLysPheHisIleIleGlyAlaValGlyIleGlyI 203

787 TTGGGGCCCTCGAGCTGCGTGCCATGATTGTGTCCATGTATCTGTACTGC 836
    ::: ||| :|||||| |||||| ||| ||| ||| |||
203 leAlaValAlaMetIlePheGlyMetIlePheSerMetIleLeuCysCys 219

837 AATCTACAA 845
    :::::: ||
220 AlaIleArg 222

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-102-942A-6

seq_documentation_block:
; Sequence 6, Application US/08102942A
; Patent No. 5726288
GENERAL INFORMATION:
  APPLICANT: Call, Katherine M.
  APPLICANT: Glaser, Thomas M.
  APPLICANT: Ito, Caryn Y.
  APPLICANT: Buckler, Alan J.
  APPLICANT: Pelletier, Jerry
  APPLICANT: Haber, Daniel A.
  APPLICANT: Rose, Elise A.
  APPLICANT: Housman, David E.
  APPLICANT: Bruening, Wendy
  APPLICANT: Darveau, Andre
  TITLE OF INVENTION: Localization and Characterization of the
  TITLE OF INVENTION: Wilms' Tumor Gene
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
    STREET: Two Militia Drive
    CITY: Lexington
    STATE: Massachusetts
    COUNTRY: U.S.
    ZIP: 02173
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/102,942A
    FILING DATE: 02-AUG-1993
    CLASSIFICATION: 530
  ATTORNEY/AGENT INFORMATION:
    NAME: Granahan, Patricia
    REGISTRATION NUMBER: 32,227

```
;   TITLE OF INVENTION:  MOLECULAR CLONING OF THE OVINE PITUITARY
;   TITLE OF INVENTION:  GROWTH HORMONE RELEASING HORMONE RECPTOR
;   NUMBER OF SEQUENCES:  18
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  POPHAM, HAIK, SCHNOBRICH & KAUFMAN, LTD.
;     STREET:  Metropolitan Square Building, Suite 800, 1450
;     STREET:  G. Street
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/660,963
;     FILING DATE:  12-JUN-1996
;     CLASSIFICATION:  536
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  O'Shaughnessy, Brian P.
;     REGISTRATION NUMBER:  32,747
;     REFERENCE/DOCKET NUMBER:  18046.036
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  202-824-8000
;     TELEFAX:  202-824-8199
;     TELEX:  248516
;   INFORMATION FOR SEQ ID NO:  12:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  498 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-660-963-12

alignment_scores:
        Quality:    98.00                         Length:       331
        Ratio:       0.790                         Gaps:          16
Percent Similarity:  37.462        Percent Identity:  23.565

alignment_block:
US-09-030-606-111 x US-08-660-963-12

Align seg 1/1  to:  US-08-660-963-12   from:  1  to: 498  .

 205  TGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTGAGAGATCTTCGGG  254
      ||||||:|||||::::||||       ||   :::|||:::||||||||
  51  TrpAlaAlaGlyCys........GlyAlaPro..........AlaSerSe   61

 255  CCACTGTCGTCCAGTGCCATGCAGTTGCAACGTGGCTACTTCCTCAT  304
      :  |||    ||||||||||:::  |||:||||||||    .:::
  61  rAlaCysTrp..AlaProCysGln...SerSerTrpAlaThrCysThrG   76

 305  CGCAGCCGCGTGTGGTCTTGGTCTTTGCTCCTTGGTTCCTGGCTGCCATGGTG  354
      |||||||:::||||||:|||||   |||     |||
  76  ln......SerValMetSerLeuSerGluArgThrSerLysHisVal   90

 355  CTAAGACTGAGAGCAAGTGCCC............  378
      ::||||    :: |||
  91  TyrLysLeuLeuLysGlyCysProThrProProTrpAlaAlaProGlySe  107

 379  ..............TCGTGACGTTCTTCTTCATCCTCCTCCTCCATCTT  412
      ::  |||      ::|||:|:::::::||:||||||
 107  rGlyThrGlyCysCysAlaGlyArgTrpGlnAlaLeuGluSerGlyAlaS  124

 413  CATTGCTGAGGTTGCAGCTGCTGTGGTGCGCCTTGG................  447
      ||   ||||||:||:  ||| ||| |||
 124  erProAlaArgLeuSerLeuThrSerAlaTrpSerGlnGlyLeuArg  140

 448  ...TGTACACCACAATGGCTGAGCACTTCCTGACGTTGCTGGTAGTGCCT  494
```

```
 234  uCysMetThrTrpAsnGlnMetAsnLeuGlyAlaThrLeuLysGlyValA  251
      |||||||     :::  |||
 664  TTGGTGACGTTGTCATTGCAACAGAATGGGGAAAGGCACTGTTCTCTTT  615
      :::::  ||  ::::||||||||:||   :::|||
 251  laAlaGlySerSerSerValLysTrpThrGluGlyGlnSerAsnHis  267
 614  GAAGTAGGG............................  606
      :::::|||
 268  SerThrGlyTyrGluSerAspAsnHisThrThrProIleLeuCysGlyAl  284
 605  ....................TGAGTCCTCAA....................  594
      |||::::::
 284  aGlnTyrArgIleHisThrHisGlyValPheArgGlyIleGlnAspValA  301
 593  ...............ATCCGTATAGTTGGTGAAGCC  573
      :::||   ::::||||:::|||
 301  rgArgValProGlyValAlaProThrLeuValArgSerAlaSerGluThr  317
 572  ACAGCACTTGAGCCCTTTCATGGTGGTGTTCCACACTTGAGTGAAGTCTT  523
      :::  ::||||| |||   :::::
 318  SerGluLysArgProPheMetCysAlaTyrProGlyCysAsnLys......  332
 522  CCTGGGAACCATAATCTTTCTTGATGGCAGGCCACTACCAGCAACGTCAGG  473
      :::||||||||||:||||:|||:::|||
 333  ....................ArgTyrPheLys..................  336
 472  AAGTGCTCAGCCATTGTGGTGTACAC........................  447
      |||||||  :::|||
 337  ....LeuSerHisLeuGlnMetHisSerArgLysHisThrGlyGluLys  351
 446  .....CAAGGCGACCACAGCAGCTGC.....AACCTCAGCAATGAAGA  409
      |||    ||| :::::||     |||:|||:::::||
 352  ProTyrGlnCysAspPheLysAspCysGluArgArgPheSerArgSerArgSerAs  368
 408  TGAGGAGGAGGATGAAGAAGAACGTCACGAGGGCCACACTTGCTCTCAGTC  359
      :::::|||||||||:|||||| ::||:||
 368  pGlnLeuLysArgHisGlnHisArgArgHisThrGlyValAlaLysLysProPheGlnLeuC  385
 358  TTAGCACCCATAGCGCCCCAGGAAACCAAGAGCAAAGACCACAACGCCCGGC  309
      ::|||||   ||||||:|| ||||||
 385  ysLysThrGlnArgLysPheSerArgSer..AspHisLeu........  398
 308  TGCCATGAGGAAGTAGCCCCACGTTGACAAACTGCAT.........GGCAC  268
      |||   |||   ||
 399  ....................LysThrHisThrArgThrHisTh  406
 267  TGGACGA............................CAGTGGCCGCGAAGATCTTC  242
      ||||:::
 406  rGlyLysThrSerGluLysProPheSerCysArgTrpProSer......  420
 241  AGAAAGGATGCCCCATCGATTGACACCCAGATGCCCACTGCCAACAGGGC  192
      |||||||
 421  ......................CysGlnLysLys  424
 191  TGCACCACAGAAAGATGAGCAAATTGAGAGGATCATCATGGTCTTAA  142
      ::::::::|||||| :::::::
 425  ...PheAlaArgSerAspGluLeuValArg......HisHisAsnMetHi  438
 141  TGAA  138
      :::
 438  sGln  439

;   seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-660-963-12
;   seq_documentation_block:
; Sequence 12, Application US/08660963
; Patent No. 5852187
;   GENERAL INFORMATION:
;     APPLICANT:  Thorner, Michael O.
;     APPLICANT:  Gaylinn, Bruce D.
;     APPLICANT:  Horikawa, Reiko
;     APPLICANT:  Lyons Jr., Charles E.
```

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
  STREET: Two Militia Drive
  CITY: Lexington
  STATE: Massachusetts
  COUNTRY: U.S.
  ZIP: 02173
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/102,942A
  FILING DATE: 02-AUG-1993
  CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
  NAME: Granahan, Patricia
  REGISTRATION NUMBER: 32,227
  REFERENCE/DOCKET NUMBER: MIT-5194A2
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617-861-6240
  TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
  LENGTH: 449 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-102-942A-4

alignment_scores:
  Quality: 97.50   Length: 456
  Ratio: 0.577   Gaps: 24
Percent Similarity: 37.061   Percent Identity: 20.175

alignment_block:
US-09-030-606-111/rev x US-08-102-942A-4 ..

Align seg 1/1 to: US-08-102-942A-4 from: 1 to: 449

1149 TCACCACTAGGCCTAGGGGGCATATCAAGGGTTTAATAGACTGGGGGAAT 1100
      || ||||||| ||||| ||||||
  16 SerSerLeuGlyGlyGlyGly..................GlyGlyCy 25

1099 GGGC.. 1096
      ||
  25 sGlyLeuProValSerGlyAlaArgGlnTrpAlaProValLeuAspPheA 42

1095 ..........AACAGAACTGGCTACCTTAGAGGCTCTGGAATGCCC 1060
                 :::: ||| ||| ||| |||
  42 laProProGlyAlaSerAlaTyrGlySerLeuGlyGlyProAlaProPro 58

1059 CCCACCCATCCA................................ 1048
      ||:::: |||
  59 ProAlaProProProProProProProHisSerPheIleLysGlnGl 75

1047 .CCCACCAATGGAGGAAGGAAAGTCAGGCATCGCTAAAAGGAGTGGTCCCTAT 999
       ||::: ||| ::|||| ||||
  75 uProSerTrpGlyGlyAlaGluProHis......GluGluGlnCysLeuS 90

998 CT........................AGCCCCAAGTCTGGAGCA 979
     || :||||||||
  90 erAlaPheThrLeuHisPheSerGlyGlnPheThrGlyThrAlaGlyAla 106

978 GAAAGGGCAGGTCCATTCTGCCCCAAGTGACATTGTTAGATCCTGTCCCC 929
     || ||||||||| ::|||
107 CysArgTyrGlyProPhe.....................GlyPr 114

928 TCCCCCAATCACTGCTTGCTGCCAGGGTGCCTCTTCACAGTTCCA..... 883
     |||||| ::: ||| ::: ::||||

```
114 oProProProSerGlnAlaSerSerGlyGlnAlaArgMetPheProAsnA 131
                           GGCTGCCGATGAGGAAGTAGC 292
882 ............TGTGGCAGCGTAGT........GGCAGA 862   :::|||::::::|||
         ||| |||:::||:::                            :::|||||||||||||:::::||
311 ....................                            410 SerGluLysProPheSerCysArgTrpHisSerCysGlnLysLysPheAl 426

131 laProTyrLeuProSerCysLeuGluSerGlnProThrIleArgAsnGln 147

                                                    291 CCACGTTGACAAACTG 276
861 GGCAGAAGTGGACTATTGTGAGATTGTAGATACAGATACATGGACAC.... 816   |::: |||:::||:
    ||| ||| |||  ::: :::::||||||                      426 aArgSerAspGluLeu 431
148 GlyTyrSerThrValThrPheAsp.GlyAlaProSerTyrGlyHisThrP 164

815 ....AATCATGGCAGCCAGCTCGAGGCCCCAATTCCAGC.......... 780   seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-571-758-4
    ::::|||::::|||||::|::  |||||| |||
164 roSerHisHisAlaAlaGlnPhe......ProAsnHisSerPheLysHis 178   seq_documentation_block:
                                                            ; Sequence 4, Application US/08571758
780 .................................................. 780   ; Patent No. 5700675
                                                            ; GENERAL INFORMATION:
179 GluAspProMetGlyGlnGlySerLeuGlyGluGlnGlnTyrSerVa    195   ;   APPLICANT: Rubin, Gerry M.
                 :::||  ||| |||||||                         ;   APPLICANT: Therrien, Marc
779 ........TGCCACCACCACCGTGACTGCATTAGTT........       749   ;   APPLICANT: Chang, Henry C.
        |||||||||         :::||                            ;   APPLICANT: Karim, Felix D.
195 lProProProValTyrGlyCysHisThrProThrAspSerCysThrGlyS 212   ;   APPLICANT: Wassarman, David A.
                                                            ;   TITLE OF INVENTION: A No. 5700675el Protein Kinase Required for Ras
748 CGGATGCTCATACAAAAGCTGATTGAAGCAACCCTCTAC.......... 711   ;   TITLE OF INVENTION: Signal Transduction
    ||  ||| |||:::  |||                                    ;   NUMBER OF SEQUENCES: 12
212 er......GlnAlaLeuLeuLeuArgThrProTyrSerSerAspAsn    225   ;   CORRESPONDENCE ADDRESS:
                                                            ;     ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
710 ...TTTTTGGTCGTGAGCCTTTG.......................     690   ;     STREET: 268 BUSH STREET, SUITE 3200
        ::: |||| ||  :::                                   ;     CITY: SAN FRANCISCO
226 LeuTyrGlnMetThrSerGlnLeuGluCysMetThrTrpAsnGlnMetAs 242   ;     STATE: CALIFORNIA
                                                            ;     COUNTRY: USA
689 .CTTGGTGCAGGTTCATTGGCTGTGTTGGTGACGTTGTCATTGCAACAG 641   ;     ZIP: 94104
     ||||||||| :::  ||| ||:::::                           ; COMPUTER READABLE FORM:
242 nLeuGlyAlaThrLeuLysGlyMetAlaAlaLysGlySerSerSerValL 259   ;   MEDIUM TYPE: Floppy disk
                                                            ;   COMPUTER: IBM PC compatible
640 AATGGGGGAAAGGC.................................... 627   ;   OPERATING SYSTEM: PC-DOS/MS-DOS
    ::||| :::||||                                          ;   SOFTWARE: PatentIn Release #1.0, Version #1.30
259 ysTrpThrGluGlyGlnSerAsnHisGlyThrGlyTyrGlySerGluAsn 275   ; CURRENT APPLICATION DATA:
                                                            ;   APPLICATION NUMBER: US/08/571,758
626 ...ACTGTTCTCTTTGAAGTAGGGTGAGTCCTCAAAATC.......... 591   ;   FILING DATE:
       ||||||||||         :::||||||                        ;   CLASSIFICATION: 435
276 HisThrAlaProIleLeuCysGlyAlaGlnTyrArgIleHisThrHisGl 292   ; ATTORNEY/AGENT INFORMATION:
                                                            ;   NAME: OSMAN, RICHARD A
590 ........................CGTATAGTTGGTGAAGCCACAG     569   ;   REGISTRATION NUMBER: 36,627
                            ||| |||   |||                   ;   REFERENCE/DOCKET NUMBER: B96-010
292 yValPheArgGlyIleGlnAspValSerGlyValSerGlyValAlaProT 309   ; TELECOMMUNICATION INFORMATION:
                                                            ;   TELEPHONE: (415) 343-4341
568 CACTT................................GAGCCCTTCATGGT 549   ;   TELEFAX: (415) 343-4342
    ::|||                                ::|||||||         ; INFORMATION FOR SEQ ID NO: 4:
309 hrLeuValArgSerAlaSerGluThrSerGlyLysArgProPheMetCys 325   ; SEQUENCE CHARACTERISTICS:
                                                            ;   LENGTH: 1003 amino acids
548 GGTGTTCCACACTTGAGTGAAGTCTTCCTGGGAACCCATAATCTTTCTTGA 499   ;   TYPE: amino acid
    :::|||||||||:::                                        ;   STRANDEDNESS: not relevant
326 AlaTyrProGlyCysAsnLys............................. 332   ;   TOPOLOGY: not relevant
                                                            ;   MOLECULE TYPE: peptide
498 TGGCAGGCACTACCAGCAACGTCAGGAAGTGCTCAGCCATTGTGGTGTAC 449   US-08-571-758-4
    :::::::|||:::  |||||||
333 ...ArgTyrPheLys..........LeuSerHisLeuGlnMetH     343   alignment_scores:
                                                            Quality:         96.50              Length:        395
448 AC........AACCTCAGCAATGAAGATGAGGAGGATGAAGAAGAACG  429   Ratio:           0.564              Gaps:          19
    ||                |||  ||| :::::|||                     Percent Similarity: 43.291      Percent Identity: 23.291
343 lsSerArgLysHisThrGlyGluLysProGlnTyrGlnCysAspPheLysAsp 359
                                                            alignment_block:
428 TGC......AACCTCAGCAATGAAGATGAGGAGGATGAAGAAGAACG    385   US-09-030-606-111 x US-08-571-758-4
    ||:::                |||  ::: :::::|||
360 CysGluArgArgPheSerArgSerArgSerAspGlnLeuLysArgHisGlnArgAr 376   Align seg 1/1 to: US-08-571-758-4  from: 1  to: 1003

384 TCACGAGGGCACACTTGCTCTCAGTCTTAGCACCATAGCACCCAG....  339       64 GAGCCTCAGCAGTTCCCTCTTTCAGAACTC...ACTGCCAAGAGCCCTGA 110
    ||| ||| |||::::  ||| ::::|||                              ||||||||||| ||:::::     ||||||| |||||||
376 gHisThrGlyValLysProPheGlnCysLysThrCysGlnIleArgLysPheS 393   322 GlnProProArgProProArgSerArgLeuProThrAspProSerProAs 338

338 .........GAAACCAAGAGCAAAGACCACAACGCC.........      312   111 ACAGAGCCACCATGCAGTGTTCAGCTTCATTAAGACCATGATGATCCT   160
    ::|||              |||::::||                              :::::||||| ||| ||:::        :::
393 erArgSerAspHisSerGlyHisLeuLysThrArgThrHisArgThrGlyLysThr 409   338 pSerHisSerSerAlaSerSerSerAspIlePheValAsp......GlyG 353
```

```
161  CTTCAATTTGCTCATCTTTCTGTGGTGCAGCCCT.............. 196
     ::||||| ||||||| :::|||  |||
353  lySerIleAsnSerSer AsnValLeuLeuValProProSerProGlyVa 369

197  ......GTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGC....... 232
         ||||||  ||||||      :::||||:::|||
369  lAlaHisValGlyMetGlyHisThrIleLysHisArgPheSerLysTrpP 386

233  ..........................................ATCCTTTCTGAA 244
                                              :::||||| :::|||
386  heGlyPheMetAlaThrCysLysLeuCysLysGlnMetSerHis 402

245  GATCTTCGGGCCACTGTCGTCCAGTGCCATGCCAGTTGTCAACGTGGGCT 294
     ::::::::||  :::||  :::|||  :::||||:::||:|
403  TrpPheLysCysThr......AspCysLysTyrIleCysHisLysSerCy 417

295  ACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGCCGTTGGTTTCCTGGGC 344
         ||||||           ||||||||
417  sAlaProHisValProProSerCysGlyLeu.................. 427

345  TGCTATGGTGCTAAGACTGAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTT 394
                                            ||||||
427  .................................................. 427

395  CATCCTCCTCCATCTTCATTG.CTGAGGTTGCAGCTGTCGTGGTCGCC 443
     ||||||    :::::::||| ::::|||          :::
428  ...ProProGluTyrValHisGluPheArgGlnThrGlnValGlyTyrAr 443

444  TTGGTGTACACCACCAATGGCTAGCACTTCCTGACGTTGCTCGGTAGTGCC 493
     ||| ::|||          ||||||
443  gTrpAspProAlaGlnHisSerSer.................. 451

494  TGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGA 543
     :::|||   ||||||||||||||||||:::|||
452  .....SerLysAlaSerProValProArgLysSerThrLeu........ 463

544  ACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCACTATACGGAT 593
                ||||:::  :::||| ::::||
464  ..........GlyLysProGlnLeuGlnLeuGlnProGlnLeuGln.. 474

594  TTTGAGGACTCAC..............CCTACTTCAAAGAGAGAAACAGTGC 628
     |||          :::::::||:::: |||||||||||||||
475  ....HisGlyAspSerSerProSerSerSerCysThrSerS 488

629  CTTTCCCCATTCTGTTGCAATG.................. 651
     |||  ||| |||:::
488  erThrProSerSerProAlaLeuPheGlnGlnGlnLeuGlnLeuAla 504

652  ACAACGTCACCACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCAC 701
     ||| :::||||  |||||||  :::||||||||||||
505  ThrProSerAlaCysGlnPro..LysProAlaPro.............. 515

702  GACCAAAAAGTAGAGGGTTGCTCAATCAGCTTTTGTATGACATCCGAAC 751
     :::|| ::: :::|||  |||:::
516  ..........AlaAlaValAlaAlaAlaAlaThrGlnGlnGln 526

752  TAATGCAGTCACCGGGTGGTGTGGCAGCTGGAATTGGGGGCCTCGAGC 801
     ::::||||| :::    :::   :: :::|| :: :: ||||
526  lyGlnGlnSerGlnPheAsnPheAsnValThrIleThrSerIleAsn 542

802  TGGCTGCCATGATTGTGTCCATGATATCTGTACTCCAATCTACAATAAGTC 851
              |||       |||  :: ||| |||:::
543  AlaCysAsnSerAsnAlaSerAlaAlaGlnThrLeuIleSerAsnGluPr 559

852  CACTTCTCGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCA 901
     :::   ||||||||| :::||||  ::::||:::
559  oGlnAlaHisMetAlaThrThrGluSerThr......LeuThrAsnGlyA 574

902  CCCTGGCAAGCAGCAGTGATTGGGGAGGGACGGGGACGGATCTAACAATGTCA 951
     ::    |||||||||||||  :::         :::    |||
574  snAsnAsnSerSerAsnAsnGlySerSerAlaAsnAsnAsnSerSer 590
```

```
952  CTTGGGCCAGAATGAGCAGCGTGCCCTTTCGTCCAGACTTGGGCTAGATA 1001
                                      |||||:::
591  ...............SerSerSerCysSerSerAsn.............. 597

1002 GGGACCACTCCTTTAGCGATGGCTGACTTTCCTTCCATTGTGGGTGGA 1051
                            ||||||  :::|||
598  .GlyHisLeu...............HisSerLeuThrGlySerG 607

1052 TGGGTGGGGGCATTCCAGAGCCTCAAGGTA................GCC 1086
     :: ||||| ||||| ::::||||| :::|||
607  lnValSerThrHisSerAlaThrSerGlnValSerAsnValSerGlySer 623

1087 AGTTCTGTTGCCCATTCCCCAGTCATTAAAC 1119
     ||||||:::::::::    :::||||::::||
624  SerSerAlaThrTyrThrSerSerLeuValAsn 634
```

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-909-984A-4

seq_documentation_block:
; Sequence 4, Application US/08909984A
; Patent No. 5747275
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wassarman, David A.
; TITLE OF INVENTION: A No. 5747275el Protein Kinase Required for Ras
; TITLE OF INVENTION: Signal Transduction
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,984A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-909-984A-4

alignment_scores:
   Quality: 96.50   Length: 395
   Ratio: 0.564   Gaps: 19
Percent Similarity: 43.291   Percent Identity: 23.291

alignment_block:
US-09-030-606-111 x US-08-909-984A-4   ..

Align seg 1/1 to:US-08-909-984A-4   from: 1 to: 1003

```
 64 GAGCCTCAGCAGTTCCCTCTTTCAGAACTC...ACTGCCAAGAGCCCTGA 110
    :::||| :::  ||| |||  |||||||::
322 GlnProProArgProArgSerArgLeuProThrAspProSerProAs 338

111 ACAGGAGCCACCATGCAGTCGTTCAGCTTCATTAAGACCATGATGATCCT 160
    |||||||||||||| :::|||:::|||           ::::
338 pSerHisSerSerAlaSerSerAspIlePheValAsp.....GlyG 353

161 CTTCAATTTGCTCATCTTCTGTGTGTGCAGCCCT............ 196
    ::||||||| ||||| ::||| |||         
353 lySerIleAsnSerSer.AsnValLeuValProProSerProGlyVa 369

197 .......GTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGC....... 232
    |||||| |||||            :::|||:::::
369 lAlaHisValGlyMetGlyHisThrIleLysHisArgPheSerLysTrpP 386

233 ...............................ATCCTTTCTGAA 244
                                  :::::|||:::
386 heGlyPheMetAlaThrCysLysLeuCysGlnLysGlnMetMetSerHis 402

245 GATCTTCGGGCCACTCGTCGTCCAGTGCCAGTTGTCAACGTGGGCT 294
    :::::::::||   :::||  ||||||||
403 TrpPheLysCysThr....AspCysLysTyrIleCysHisLysSerCy 417

295 ACTTCCTCATCGCAGCCGGCGGGTTGTGGTCTTTGCTCTTGGTTTCCTGGGGC 344
    ||||||
417 sAlaProHisValProProSerCysGlyLeu................. 427

345 TGCTATGGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTT 394
                                     
427 ................................................. 427

395 CATCCTCCTCCTCATCTTCATTG..CTGAGGTTGCAGCTGCTGTGGTCGCC 443
    ||||||  :::::||| ::|||  :::
428 ...ProProGluTyrValHisGluPheArgGlnThrGlnValGlyIleTyrAr 443

444 TTGGTGTACACCACCATGGCTGAGCACTTCCTGACGTTGCTGGTAGTGCC 493
    |||     :::  |||
443 gTrpAspProAlaGlnHisSerSer..................... 451

494 TGCCATCAAGAAGAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGA 543
    |||::   ||||||||||||:::::|||
452 ....SerLysAlaSerProValProArgLysSerThrLeu......... 463

544 ACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCCACCACTATACGGAT 593
    ||| |||  ||||||||||  :::||
464 ..........GlyLysProGlnLeuGlnGlnGlnLeuGln.. 474

594 TTTGAGGACTCAC..........CCTACTTCAAAAGAGAACAGTGC 628
    |||::    |||  |||:::::|||:::  |||
475 ........HisGlyAspSerSerProSerSerCysThrSerS 488

629 CTTTCCCCATTCTGTTGCAATG............................. 651
    |||  |||  ||:::::
488 erThrProSerProAlaLeuPheGlnGlnGlnGlnLeuGlnLeuUAla 504

652 ACAACGTCACCAACACAGCCAATGAAACCTGCACCAAGCAAAGGCTCAC 701
    ::|||||| |||  ::::::  |||  ::::::
505 ThrProSerAlaCysGlnPro...LysProAlaPro........... 515

702 GACCAAAAAGTAGAGGGTTGCTTCAATCAGCTTTGTATGACATCCGAAC 751
                 ||||||| :::  |||:::|||
516 ...........AlaAlaValAlaAlaAlaAlaGlnThrGlnInG 526

752 TAATGCAGTCACCGTGGGTGTGTGGCAGCTGGAATTGGGGCCTCGAGC 801
    ::||||||  :::    ::::::  :::  :::
526 lyGlnGlnSerGlnPheAsnPheProAsnValThrIleThrSerIleAsn 542

802 TGGCTGCCATGATTGTGTCCATGATATCTGACTGCAATCCAATAAGTC 851
    :::    |||:::  ||| :::::  
543 AlaCysAsnAsnAlaSerAlaAlaGlnThrLeuIleLeuSerAsnGluPr 559

852 CACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGCA 901
```

```
559 oGlnAlaHisMetAlaThrThrGluSerThr.....LeuThrAsnGlyA 574
    :::    |||||||||  :::|||      ||| :::|||:
 64

902 CCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCA 951
    ||||||||||::: |||     ::: :::  |||
574 snAsnAsnSerSerAsnAsnGlySerAlaAsnAsnSerSer 590

952 CTTGGGCCAGAATGGACCTGCCCTTCTGCTCCAGACTTGGGCCTAGATA 1001
    :::   ||||||:::
591 ............SerSerSerCysSerAsn............. 597

1002 GGGACCACTCCTTTTAGCGATGCCGATCTTCCTTCCATTGTGGGTGGA 1051
     |||::: |||          ||||||:: ||| |||
598 .GlyHisLeu..............HisSerLeuThrGlySerG 607

1052 TGGGTGGGGGCATTCCAGAGCCTCTAAGGTA...........GCC 1086
     ::: ||||| :::|||:::::||
607 lnValSerThrHisSerAlaThrSerGlnValSerAsnValSerGlySer 623

1087 AGTTCTGTTGCCCATTCCCCCAGTCTATTAAAC 1119
     ||||||:::::||:::::||| ||||||:::|||
624 SerSerAlaThrTyrThrSerLeuValAsn 634
```

```
                        GenCore version 4.5
                  Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        September 28, 1999, 12:21:27 ; Search time 2928.69 Seconds
                                     (without alignments)
                                     1399.745 Million cell updates/sec

Title:         US-09-030-606-111
Perfect score: 1289
Sequence       1 AGCCAGGCGTCCCTCTGCCT..........GTTAAAAAAAAAAAAAAAA 1289

Scoring table: IDENTITY_NUC

Searched:      679419 seqs, 159015680 residues

Database :     GenEmbl:*
               1: gb_bal:*
               2: gb_ba2:*
               3: gb_om:*
               4: gb_ov:*
               5: gb_pat:*
               6: gb_ph:*
               7: gb_pl1:*
               8: gb_pl2:*
               9: gb_prl:*
              10: gb_pr2:*
              11: gb_pr3:*
              12: gb_ro:*
              13: gb_st:*
              14: gb_sts:*
              15: gb_sy:*
              16: gb_un:*
              17: gb_vi:*
              18: em_fun:*
              19: em_htg:*
              20: em_hum1:*
              21: em_hum2:*
              22: em_in:*
              23: em_om:*
              24: em_or:*
              25: em_ov:*
              26: em_pat:*
              27: em_ph:*
              28: em_pl:*
              29: em_ro:*
              30: em_sts:*
              31: em_sy:*
              32: em_un:*
              33: em_vi:*
              34: gb_htg1:*
              35: gb_htg2:*
              36: gb_in1:*
              37: gb_in2:*
              38: em_bal:*
              39: em_ba2:*
              40: em_hum3:*
              41: em_hum4:*
              42: gb_pr4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1266	98.2	1278	11	AF065388	AF065388 Homo sapi

```
 2  1064.8  82.6  1076  11  AF054838   AF054838 Homo sapi
 3    84.6   6.6   849  11  AF054841   AF054841 Homo sapi
 4    83     6.4  1358  11  AF022813   AF022813 Homo sapi
 5    81.8   6.3  1432  11  AF089749   AF089749 Homo sapi
 6    73.2   5.7   835  12  AF052492   AF052492 Mus muscu
 7    72.2   5.6  1496   9  HUMTAPA1   M33680 Human 26-kD
 8    71.6   5.6  1720  12  MUSTALLA   D26483 Mouse mRNA
 9    70.6   5.5   711  11  AF116600   AF116600 Pan trogl
10    67.8   5.3   711  11  AF116599   AF116599 Chloroceb
11    62.8   4.9  1452  12  HUMCD53    M60871 Human cell
12    62.8   4.9  1480   9  HUMCD53GLY M37033 Human CD53
13    62.2   4.8  1303  12  RNU19894   U19894 Rattus norv
14    60.4   4.7  1743   9  HUMA15     D10653 Homo sapien
15    60.4   4.7  1792   9  HUMMRNAB   L10373 Human (clon
16    60.4   4.7  1731   9  HUMTALLA1  D29808 Human mRNA
17    60.4   4.7  1234  12  MMMD3MRN   X59047 M.musculus
18    59     4.6  1699  12  RATOX44    M57276 Rat leukocy
19    57.6   4.5  1657  12  MUSC33R2IA D14883 Mouse mRNA
20    56     4.3   687   5  AR016440   AR016440 Sequence
21    56     4.3  1120   5  AR016441   AR016441 Sequence
22    56     4.3   687   5  E05732     E05732 cDNA encodi
23    56     4.3   687   5  I13743     I13743 Sequence 2
24    56     4.3  1120   5  I13744     I13744 Sequence 3
25    56     4.3  1120   9  AGMDRAP27  D10726 African gre
26    56     4.3   687   9  HUMP24CD9A L34068 Human antig
27    56     4.3  1120  10  HSMRP1     X60111 H.sapiens m
28    56     4.3  1192  10  HUMANTCD9  M38690 Human CD9 a
29    55.4   4.3   875   9  HUMCD63    M58485 Human lysos
30    55.4   4.3   850   9  HUMOGCB    M59907 Human melan
31    55.4   4.3   826  10  HSME491    X07982 Human mRNA
32    55.4   4.3  2049  10  HSR2IMP    X53795 Human R2 mR
33    55.4   4.3  1624  10  S48196     S48196 C33 antigen
34    55.4   4.3   818  10  S93788     S93788 ocular mela
35    55.4   4.3  1607  11  HSU20770   U20770 Human metas
36    55     4.3   854   3  RABME491CD D21264 Rabbit mRNA
37    55     4.3  1130  12  MUSCD9ANT  L08115 Mus musculu
38    54.2   4.2   660  12  MMCSGCD53  X97227 M.musculus
39    54.2   4.2   612  12  MMIGCD53A  Z16071 Mus musculu
40    53.6   4.1 38397  36  DMCD8B     AL022018 Drosophil
41    53.4   4.1  1740  12  AF049882   AF049882 Rattus no
42    53.4   4.1 41230  36  CELC14A11  U97592 Caenorhabdi
43    52.8   4.1   669  11  AF054839   AF054839 Homo sapi
44    51.6   4.0  1819  12  D89290     D89290 Mus musculu
45    51.6   4.0  1641  12  MMU89772   U89772 Mus musculu
```

ALIGNMENTS

```
RESULT   1
AF065388
LOCUS       AF065388    1278 bp    mRNA            PRI       25-MAY-1998
DEFINITION  Homo sapiens tetraspan NET-1 mRNA, complete cds.
ACCESSION   AF065388
NID         g3152700
VERSION     AF065388.1  GI:3152700
KEYWORDS    .
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1278)
AUTHORS     Rubinstein,E., Serru,V. and Boucheix,C.
TITLE       New tetraspans identified in the EST database
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1278)
AUTHORS     Rubinstein,E., Serru,V. and Boucheix,C.
TITLE       Direct Submission
JOURNAL     Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier,
            Villejuif 94807, France
FEATURES             Location/Qualifiers
     source          1..1278
                     /organism="Homo sapiens"
```

```
CDS              /db_xref="taxon:9606"
                 /clone="IMAGE ID 511239"
           122. .847
                 /note="similar to Tspan-1; TM4SF"
                 /codon_start=1
                 /product="tetraspan NET-1"
                 /protein_id="AAC17119.1"
                 /db_xref="PID:g3152701"
                 /db_xref="GI:3152701"
                 /translation="MQCFSFIKTMMILFNLLIFLCGAALLAVGIWVSIDGASFLKIFG
                 PLSSSAMQFVNVGYFLIAAGVVFALGFLGCYGAKTESKCALVTFFILLLIFIAEVA
                 AVVALVYTIMAEHFLILLVVPAIKKDYGSOEDFTQVWNTIMKGLKCCGFTNYTDFED
                 SPYFKENSAFPPFCCNDNVTNTANETCTKQKAHDQKVEGCFNQLLYDIRTNAVTVGGV
                 AAGIGGLELAAMIVSMYLYCNLQ"

BASE COUNT       280 a      335 c      329 g      334 t
ORIGIN

Query Match      98.2%;   Score 1266;   DB 11;   Length 1278;
Best Local Similarity  99.9%;   Pred. No. 0;
Matches 1277;   Conservative  0;   Mismatches  0;   Indels  1;   Gaps  1;

Qy     2 GCCAGGCGCTCCCTCGCCTGCCCACTCAGTGCCAACACCCGGAGCTGTTTTGTCCTTTG   61
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GCCAGGCGTCCCTCGCCTGCCCACTCAGTGCCAACACCCGGAGCTGTTTTGTCCTTTG   60

Qy    62 TGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCAC  121
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCAC  120

Qy   122 CATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCT  181
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCT  180

Qy   182 GTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCT  241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCT  240

Qy   242 GAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCCAGTTGTCAAGCTGGGCTACTTCCT  301
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCCAGTTGTCAAGCTGGGCTACTTCCT  300

Qy   302 CATCGCAGCCGGCGTTGGGTCTTGGTCTTGGTTCCTGGGCGTGCTATGGTGCTAAGAC  361
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CATCGCAGCCGGCGTTGGGTCTTGGTCTTGGTTCCTGGGCGTGCTATGGTGCTAAGAC  360

Qy   362 TGAGAGCAAGTGCCCCGTGACGTTCTTCTTCATCCTCCTCCTCCTCATCTTCATTGCTGA  421
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGAGAGCAAGTGCCCCGTGACGTTCTTCTTCATCCTCCTCCTCCTCATCTTCATTGCTGA  420

Qy   422 GGTTGCAGCTGCTGTGGCGTTGGTCGCCCTGGTGTACACACATGCCTTCACTATGGTTG  481
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGTTGCAGCTGCTGTGGCGTTGGTCGCCCTGGTGTACACACATGCCTTCACTATGGTTG  480

Qy   482 GCTGGTAGTGCCTGCCATCAAGAAGAATTATGGTTCCCAGGAAGACTTCACTCAAGTGTG  541
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GCTGGTAGTGCCTGCCATCAAGAAGAATTATGGTTCCCAGGAAGACTTCACTCAAGTGTG  540

Qy   542 GAACACCACCATGAAAAGGGCTCAAGTGCTTGTGCTTCACCAACTATACGGATTTTGAGGA  601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAACACCACCATGAAAAGGGCTCAAGTGCTTGTGCTTCACCAACTATACGGATTTTGAGGA  600

Qy   602 CTCACCCTACTTCAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCAC  661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CTCACCCTACTTCAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCAC  660

Qy   662 CAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTG  721
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTTG  720

Qy   722 CTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGTGTGGCAGC  781
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 .CTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGTGTGGCAGC  780
```

```
Qy   782 TGGAATTGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGATCTGTACTGCAATCT  841
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TGGAATTGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGATCTGTACTGCAATCT  840

Qy   842 ACAATAAGTCCACTTCTGCCTCTGCCCACTACTGCTGCTGCCACATGTGAAGAGGCA  901
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ACAATAAGTCCACTTCCGCCTCTGCCCACTACTGCTGCTGCCACATGTGAAGAGGCA  900

Qy   902 CCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCAG  961
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCAG  960

Qy   962 AATGGACCTGCCCTTTCTGCTCCAGACTTGGGGCTGTAGATAGGGACCATTCCAGAGCCTCTAAG-GCG  1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   961 AATGGACCTGCCCTTTCTGCTCCAGACTTGGGGCTGTAGATAGGGACCACTCCTTTTAGGCG  1020

Qy  1021 ATGCCTGACTTTCCTTCCATTGTGGGTGGATGGGTCGGGGGGGCATTCCAGAGCCTCTAAG  1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATGCCTGACTTTCCTTCCATTGTGGGTGGATGGGTCGGGGGGGCATTCCAGAGCCTCTAAG  1080

Qy  1081 GTAGCCAGTTCGTTGCCCATTCCCCAGTCTTATTAATAACCCTTGATATGCCCCTAGGCC  1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GTAGCCAGTTCGTTGCCCATTCCCCAGTCTTATTAATAACCCTTGATATGCCCCTAGGCC  1140

Qy  1141 TAGTGGTGATCCCAGTGCTCTACTGGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT  1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 TAGTGGTGATCCCAGTGCTCTACTGGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT  1200

Qy  1201 AAGTGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAAGGCACTTCAAAATGCATAAACC  1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AAGTGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAAGGCACTTCAAAATGCATAAACC  1260

Qy  1261 TGTTACAATGTTAAAAAA  1278
         ||||||||||||||||||
Db  1261 TGTTACAATGTTAAAAAA  1278
```

```
RESULT   2
AF054838
LOCUS         AF054838        1076 bp        mRNA                 PRI        03-NOV-1998
DEFINITION    Homo sapiens tetraspan TM4SF (TSPAN-1) mRNA, complete cds.
ACCESSION     AF054838
NID           g2997740
VERSION       AF054838.1        GI:2997740
KEYWORDS      .
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 1076)
  AUTHORS     Todd,S.C., Doctor,V.S. and Levy,S.
  TITLE       Sequences and expression of six new members of the
              tetraspanin/TM4SF family
  JOURNAL     Biochim. Biophys. Acta 1399 (1), 101-104 (1998)
  MEDLINE     98390278
REFERENCE     2  (bases 1 to 1076)
  AUTHORS     Todd,S.C., Doctor,V.S. and Levy,S.
  TITLE       Direct Submission
  JOURNAL     Submitted (19-MAR-1998) Medicine, Stanford, 300 Pasteur, Stanford,
              CA 94305-5115, USA
FEATURES             Location/Qualifiers
     source          1..1076
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..1076
                     /gene="TSPAN-1"
     CDS             1..726
                     /gene="TSPAN-1"
                     /note="tetraspanin; transmembrane 4 super family; Tspan-1"
                     /codon_start=1
                     /product="tetraspan TM4SF"
                     /protein_id="AAC69714.1"
```

```
                    /db_xref="PID:g2997741"                                              Db   841 ATGGACCTGCCTTTCTGCTCCAGACTTGGGCTAGATAGGGACCACTCCTTTAGGCGA  900
                    /db_xref="GI:2997741"                                                     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                    /translation="MQCFSFIKTMMILFNLLIFLCGAALLAVGIWVSIDGASFLKIFG             Qy  1023 GCCTGACTTTCCTTCCATTGGTGGGTGGATGGGTGGGGGCATTCCAGAGCCTCTAAGGT 1082
                    PLSSSAMQFFVNVGYFLIAAGVVVFALGFLGCYGAKTESKCALVTFFFILLLIFIAEVA                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                    AAVVALVYTMAEHFLTLLVVPAIKKDYGSQEDFTQVWNTMKGLKCCGFTNYTDFED               Db   901 TGCTGACTTTCCTTCCATTGGTGGGTGGATGGGTGGGGGCATTCCAGAGCCTCTAAGGT  960
                    SPYFKENSAFPPFCCNDNVTNTANETCTEQKAHDQKVEGCFNQLLYDIRTNAVTVGGV
                    AAGIGGLELAAMIVSMYLYCNLQ"                                             Qy  1083 AGCCAGTTCGTTGCCAGTCGTTGCCATTCCCCAGTCTTATTAAACCCTGATATGCCCCTAGGCCTA 1142
                                                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  BASE COUNT   226 a    278 c    283 g    289 t                              Db   961 AGCCAGTTCGTTGCCAGTCGTTGCCATTCCCCAGTCTTATTAAACCCTTGATATGCCCCTAGGCCTA 1020

  Query Match             82.6%;   Score 1064.8;  DB 11;  Length 1076;        Qy  1143 GTGGTGATCCAGTGCTCTACTGGGGGATGAGAAAGGCATTTATAGGCCTGGGC 1198
  Best Local Similarity   99.3%;   Pred. No. 4.2e-298;                               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
  Matches 1069;   Conservative    0;   Mismatches    0;   Indels    7;  Gaps  0;  Db  1021 GTGGTGATCCAGTGCTCTACTGGGGGATGAGAGAAAGGCATTTATAGGCCTGGGC 1076

  Qy   123 ATGCAGTGCTTCAGCTTCATTAAGACAATGATGATCCTTCAATTTGCTCATCTTTCTG  182     RESULT 3
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||      AF054841
  Db     1 ATGCAGTGCTTCAGCTTCATTAAGACAATGATGATCCTTCAATTTGCTCATCTTTCTG   60     LOCUS       AF054841     849 bp    mRNA              PRI       03-NOV-1998
                                                                              DEFINITION  Homo sapiens tetraspan TM4SF (TSPAN-4) mRNA, complete cds.
  Qy   183 TGTGGTGCAGCCCTGGCAGTGGGCATCGGGTCAATCGATGGGGCATCCTTTCTG  242     ACCESSION   AF054841
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||           NID         g2997746
  Db    61 TGTGGTGCAGCCCTGGCAGTGGGCATCGGGTCAATCGATGGGGCATCCTTTCTG  120     VERSION     AF054841.1   GI:2997746
                                                                              KEYWORDS    .
  Qy   243 AAGATCTTCGGGCCACTGTCGTCAGTGCCATGCAGTTTGTCAACGTGGGCTACTTCCTC  302     SOURCE      human.
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||        ORGANISM   Homo sapiens
  Db   121 AAGATCTTCGGGCCACTGTCGTCAGTGCCATGCAGTTTGTCAACGTGGGCTACTTCCTC  180                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                                                                                          Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Qy   303 ATCGCCGCCGGCGTTGTGGCTTCTGCTCTGGGCTGCTATGGTGCTAAGACT  362     REFERENCE   1 (bases 1 to 849)
            |||||||||||||||||||||||||||||||||||||||||||||||||||          AUTHORS     Todd,S.C. Doctor,V.S. and Levy,S.
  Db   181 ATCGCCGCCGGCGTTGTGGCTTCTGCTCTGGGCTGCTATGGTGCTAAGACT  240     TITLE       Sequences and expression of six new members of the
                                                                                          tetraspanin/TM4SF family
  Qy   363 GAGAGCAAGTGCGCCCTGACGTTCTTCTTCATCCTCCTCCTCTTCATTGCTGAG  422     JOURNAL     Biochim. Biophys. Acta 1399 (1), 101-104 (1998)
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||            MEDLINE     98390278
  Db   241 GAGAGCAAGTGCGCCCTGACGTTCTTCTTCATCCTCCTCCTCTTCATTGCTGAG  300     REFERENCE   2 (bases 1 to 849)
                                                                              AUTHORS     Todd,S.C. Doctor,V.S. and Levy,S.
  Qy   423 GTTGCAGCTGCTGTGTCGGCCTTGGTGTACACCACCAAATGGCTGAGCACTTCCTGACGTTG  482     TITLE       Direct Submission
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||      JOURNAL     Submitted (19-MAR-1998) Medicine, Stanford, 300 Pasteur, Stanford,
  Db   301 GTTGCAGCTGCTGTGTCGGCCTTGGTGTACACCACCAAATGGCTGAGCACTTCCTGACGTTG  360                 CA 94305-5115, USA
                                                                              FEATURES    Location/Qualifiers
  Qy   483 CTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGAGACTTCACTCAAGTGTGG  542     source      1..849
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||                 /organism="Homo sapiens"
  Db   361 CTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGAGACTTCACTCAAGTGTGG  420                 /db_xref="taxon:9606"
                                                                              gene        1..849
  Qy   543 AACACCACCATGAAAGGCTCAAGTGCTGTGGCTTCACCGGATTTTGAGGAC  602                 /gene="TSPAN-4"
            ||||||||||||||||||||||||||||||||||||||||||||||||||              CDS         119..835
  Db   421 AACACCACCATGAAAGGCTCAAGTGCTGTGGCTTCACCGGATTTTGAGGAC  480                 /gene="TSPAN-4"
                                                                                          /note="tetraspanin; transmembrane 4 super family; NAG-2;
  Qy   603 TCACCCTACTTCAAAGAGAACAGTGCCCCATTCGTTGTTGCAATGACAACGTCACC  662                 Tspan-4"
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||            /codon_start=1
  Db   481 TCACCCTACTTCAAAGAGAACAGTGCCCCATTCGTTGTTGCAATGACAACGTCACC  540                 /product="tetraspan TM4SF"
                                                                                          /protein_id="AAC69717.1"
  Qy   663 AACACAGCCAATGAAACCTGCACCGAGCAAAAGGCTCACGGGTTGC  722                 /db_xref="PID:g2997747"
            ||||||||||||||||||||||||||||||||||||||||||||||                  /db_xref="GI:2997747"
  Db   541 AACACAGCCAATGAAACCTGCACCGAGCAAAAGGCTCACGGGTTGC  600                 /translation="MARACLQAVKYLMFAFNLLFWLGGCGVLGVGIWLAATQGSFATL
                                                                                          SSSFPSLSAANLLIITGAFVMAIGFVGCLGAIKENKCLLLTFFLLLLVFLLEAIIAI
  Qy   723 TTCAATGCTTTTGTATGACATCCGAACTAATGCAGTCGTGGTGTGCAGCT  782                 LFFAYTDKIDRYAQQDLKKGLHLYGTQGNVGLTNAWSIIQTDFRCCGVSNYTDWFEVY
            ||||||||||||||||||||||||||||||||||||||||||||||||||                  NATRVPDSCCLEFSESCGLHAPGTWWKAPCYETVKVWLQENLLAVGIFGLCTALVQIL
  Db   601 TTCAATGCTTTTGTATGACATCCGAACTAATGCAGTCGTGGTGTGCAGCT  660                 GLTFAMTMYCQVVKADTYCA"

  Qy   783 GGAATTGGGGCCCTGCAGCTGGCCATGATTGTGTCCATGTACTGCAATCTA  842     BASE COUNT   132 a    284 c    249 g    184 t
            ||||||||||||||||||||||||||||||||||||||||||||||||||              ORIGIN
  Db   661 GGAATTGGGGCCCTGCAGCTGGCCATGATTGTGTCCATGTACTGCAATCTA  720
                                                                              Query Match             6.6%;    Score 84.6;  DB 11;  Length 849;
  Qy   843 CAATAAGTCCACTCTGCCCTCGCCTGCTACCTGCTGCCACATGGGAACGTGAAGAGGCAC  902     Best Local Similarity   53.7%;   Pred. No. 5.8e-14;
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||       Matches 252;   Conservative    0;   Mismatches 199;   Indels  18;  Gaps  3;
  Db   721 CAATAAGTCCACTCTGCCCTCGCCTGCTACCTGCTGCCACATGGGAACGTGAAGAGGCAC  780
                                                                              Qy   129 TGCTTCAGCTTCATTAAGACAACCATGATGATCCTCTTCAATTTGCTCATCTTTCTGTGTGGT  188
  Qy   903 CCTGCAAGCAAGCAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCAGA  962                 | |||||| |||||||||| ||||||  |||||| || |||||||||||| |||| ||
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||       Db   131 TGCCTCCAGGCCGTCAAGTACCTCATGTTCGCCTTCAACCTGCTCTTCTGGCTGGGAGGC  190
  Db   781 CCTGCAAGCAAGCAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCAGA  840
                                                                              Qy   189 GCAGCCCTGTTGGCAGTGGGCATCGGGTCATGCTGGTCATCGATGGGGCATCCTTTCTGAGATC  248
  Qy   963 ATGGACCTGCCTTTCTGCTCCAGACTTGGGCTAGATAGGGACCACTCCTTTAGGCGAT 1022             | |||| || |||| | |||||||||| |||| |||||  | |||||||| |    ||
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||      Db   191 TGTGGGCTGCTGGTCTGGGTGTCGGCATCTGGCT----------GGCCGCCACAGGGGAGC  238
```

```
Query Match                        6.4%;    Score 83;    DB 11;    Length 1358;
Best Local Similarity             53.5%;    Pred. No. 1.8e-13;
Matches  251;   Conservative   0;   Mismatches 200;   Indels  18;   Gaps   3;

Qy    129  TGCTTCAGCTTCATTAAGACCATGATGATCCTTCATTTGCTCATCTTTCTGTGTGGT  188
                |||||  |||  ||  |||  ||||||||  | |||| ||||||| ||| ||||||
Db    117  TGCCTCCAGGCCGTCAAGTACCTCATGTTCGCCTTCAACCTGCTCTTCGGCTGGGAGGC  176

Qy    189  GCAGCCCGTTGCAGTGGCATCTGGGTGTCAATGCAGTGTCAATGATGGGCATCCTTTCTGAAGATC  248
                | |  | |||  |||||   ||||  | |||||   |||   ||| | | |
Db    177  TGTGGCGGTGCTGGGTGTCGGCATCTGGCT-------GCCGCCACACAGGGAGC  224

Qy    249  TTCGGGCCACTGCTGCGTCCAGTGCCATGCCAGTTGTCAACGTGGGCTACTTCCTCATCGCA  308
                ||||  |||||| |||||| ||    |  || | |||   ||||| | |||||||| ||
Db    225  TTCGCCACGCTGCCTC---TTCCTTCCGTCCGTGTGGGCTGTGCTCCAACCTGCTCATCATC  281

Qy    309  GCCGGCGGTTGTGGTCTTTGCTCTGGCTGCTATGGTGCTAAGACTGAGAGC  368
                ||||| | || | || ||||| |||  |||||  ||||| ||||||||||
Db    282  ACCGGCGGCCTTTGTCATGGCCATCGGCTTCGTGGGCTGCCATCAAGGAGAAC  341

Qy    369  AAGTGTGCCCGTGACGTTCTTTCATCCTCCTCCTCATTGCTGAGGTTGCA  428
                ||||| | |||| |  ||||| |||||||||| ||||||| |||||||||
Db    342  AAGTGCCTCCTGCTCACTTTCTTCCTGCTGCGTGGTGTTCCTGTCGAGGTTGCA  401

Qy    429  GCTGCTGTGTGGTCGCCTTGGTGTACAC---CACAATGGCTGAGCACTTCCTGACGTTGCTG  485
                ||  | ||  |   || ||||| |||||       || ||||| ||||| ||||||  |||
Db    402  ATCGCCATCCTCTTCTTCCGCCTACACGGACAAGATTGACAGGTATGCCCAGCAAGACCTG  461

Qy    486  GTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAAC  545
                || || ||| |||||||||||||  || ||| |||| ||| |||| | | |||||||||
Db    462  AAGAAAGGTTGCACCTGTACGGCACGCAGGGCAACGTGGGCCTCACCAACGCCTGGAGC  521

Qy    546  ACCACCATGAAAGGGCTCAAGTGCTGTGCCTTCACCAACTATACGGATT  594
                ||||| ||||||||| ||||   || || | | ||||| |  | ||||
Db    522  ATCATCCAGACCGACCTTCCGCTGCTGTGGCGTCTCCAACTACACTGACT  570


RESULT  5
AF089749
LOCUS        AF089749        1432 bp        mRNA                    PRI        16-SEP-1998
DEFINITION   Homo sapiens tetraspan NET-5 mRNA, complete cds.
ACCESSION    AF089749
NID          g3603458
VERSION      AF089749.1        GI:3603458
KEYWORDS     .
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1432)
AUTHORS      Rubinstein,E., Serru,V., Dessen,P. and Boucheix,C.
TITLE        New tetraspans identified in the EST database
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 1432)
AUTHORS      Rubinstein,E., Serru,V., Dessen,P. and Boucheix,C.
TITLE        Direct Submission
JOURNAL      Submitted (01-SEP-1998) INSERM U268, 14 av Paul Vaillant Couturier,
             Villejuif 94807, France
FEATURES             Location/Qualifiers
     source          1..1432
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             127..846
                     /note="TM4SF; transmembrane 4 super family"
                     /codon_start=1
                     /product="tetraspan NET-5"
                     /protein_id="AAC35859.1"
                     /db_xref="PID:g3603459"
                     /db_xref="GI:3603459"
                     /translation="MARGCLCLCLKYMMFLFNLIFWLCGCLLGVGIWLSVSQGNFATF
                     SPSFPSLSAANLVIAIGITVMVTGFLGCLGAIKENKCLLLSFFIVLLLAEGILLI
```

Left column (rotated):

```
Qy    249  TTCGGGCCACTGCTGCGTCCAGTGCCATGCCAGTTGTCAACGTGGGCTACTTCCTCATCGCA  308
                |||||  |||||| |||||| ||    |  || | |||   ||||| | |||||||| ||
Db    239  TTCGCCACGCTGCGTCCTC---TTCCTTCCGTCCGTGTGGGCTGTGCTCCAACCTGCTCATC  295

Qy    309  GCCGGCGGTTGTGGTCTTTGCTCTGGCTGCTATGGTGCTAAGACTGAGAGC  368
                ||||| | || | || ||||| |||  |||||  ||||| ||||||||||
Db    296  ACCGGCGGCCTTTGTCATGGCCATCGGCTTCGTGGGCTGCCATCAAGGAGAAC  355

Qy    369  AAGTGTGCCCGTGACGTTCTTTCATCCTCCTCCTCATTGCTGAGGTTGCA  428
                ||||| | |||| |  ||||| |||||||||| ||||||| |||||||||
Db    356  AAGTGCCTCCTGCTCACTTTCTTCCTGCTGCGTGGTGTTCCTGTCGAGGTTGCA  415

Qy    429  GCTGCTGTGTGGTCGCCTTGGTGTACAC---CACAATGGCTGTGCCTTCACCAACTATACGGATT  485
                ||  | ||  |   || ||||| |||||       || |||| |  |||| ||||||  ||
Db    416  ATCGCCATCCTCTTCTTCCGCCTACACGGACAAGATTGACAGGTATGCCCAGCAAGACCTG  475

Qy    486  GTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAAC  545
                || || ||| |||||||||||||  || ||| |||| ||| |||| | | |||||||||
Db    476  AAGAAAGGTTGCACCTGTACGGCACGCAGGGCAACGTGGGCCTCACCAACGCCTGGAGC  535

Qy    546  ACCACCATGAAAGGGCTCAAGTGCTGTGCCTTCACCAACTATACGGATT  594
                ||||| ||||||||| ||||   || || | | ||||| |  | ||||
Db    536  ATCATCCAGACCGACCTTCCGCTGCTGTGGCGTCTCCAACTACACTGACT  584


RESULT  4
AF022813
LOCUS        AF022813        1358 bp        mRNA                    PRI        18-NOV-1997
DEFINITION   Homo sapiens tetraspan (NAG-2) mRNA, complete cds.
ACCESSION    AF022813
NID          g2586349
VERSION      AF022813.1        GI:2586349
KEYWORDS     .
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1358)
AUTHORS      Tachibana,I., Bodorova,J., Berditchevski,F., Zutter,M.M. and
             Hemler,M.E.
TITLE        NAG-2, a novel transmembrane-4 superfamily (TM4SF) protein that
             complexes with integrins and other TM4SF proteins
JOURNAL      J. Biol. Chem. 272 (46), 29181-29189 (1997)
MEDLINE      98030601
REFERENCE    2  (bases 1 to 1358)
AUTHORS      Tachibana,I., Bodorova,J., Berditchevski,F., Zutter,M.M. and
             Hemler,M.E.
TITLE        Direct Submission
JOURNAL      Submitted (04-SEP-1997) Tumor Virology, Dana-Farber Cancer
             Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES             Location/Qualifiers
     source          1..1358
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="breast cancer"
                     /cell_line="MDA-MB-435"
     gene            1..1358
                     /gene="NAG-2"
     CDS             105..821
                     /gene="NAG-2"
                     /codon_start=1
                     /product="tetraspan"
                     /protein_id="AAC51864.1"
                     /db_xref="PID:g2586350"
                     /db_xref="GI:2586350"
                     /translation="MARACLQAVKYLMFAFNLLFWLGGCGVLGVGIWLAATGGSFATL
                     SSSFPSLSAANLIITGAFVMAIGFVGCLGAIKENKCLLLTFFLLLLVFLLEATIAI
                     LFFAYTDKIDRYAQQDLKKGLHLYGTQGNVGLTNAWSIIQTDFRCCGVSNYTDWFEVY
                     NATRVPDSCCLEFSESCCLHAPGTWWKAPCYETVKVWKIQENLLAVGIFGLCTALVQIL
                     GLIFAMTMYCQVVKADIYCA"
BASE COUNT    265  a    428  c    400  g    265  t
ORIGIN
```

```
                    LFFVYMDKVNENAKKDLKEGLLLYHTENNVGLKNAWNIIQAEMRCCGVTDYTDWYPVL
                    GENTVPDRCCMENSQGGCRNATTPLWRTGCYEKVKMWFDDNKHVLGTVGMCILIMQIL
                    GMAFSMTLFQHIHRTGKKYDA"
BASE COUNT    296 a    410 c    405 g    321 t
ORIGIN

Query Match               6.3%;   Score 81.8;  DB 11;  Length 1432;
Best Local Similarity    51.8%;   Pred. No. 4.1e-13;
Matches  247;  Conservative    0;  Mismatches  212;  Indels   18;  Gaps    2;

Qy    121  CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC  180
           | |||| | ||| || ||| ||| | || || |||||||||||| |  ||  ||  |||
Db    131  CCAGGGGCTGCCCTGCGTCCTTGAAGTACATGATGATGTTCCTCTTCAATTGATATTCTGGC  190

Qy    181  TGTGTGGGTGCAGGCCCTGTTGGCAGTGGGCATCTGGTGTCAATCGATGGGCATCCTTTC  240
           | ||||| || ||| ||  | | |||| ||| || ||  |  |  ||| |
Db    191  TCTGTGGGTGTGGGCTGCTGGGGAGTGGGCATCTGGCCTCTCCGTGTCCAAGGCAACTTTG  250

Qy    241  TGAAGATCTTCGGGCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC  300
           ||               ||| ||| | ||  || | ||| |
Db    251  CCACCTTC--------------TCCCCAGTTCCCTTCGTTGTCTGCGACCAACCTGG  295

Qy    301  TCATCGCAGCCGGCGTTGTGGTCTTTGCTCCGGGTCCTGCTATGGTGCTAAGA  360
           |||| |   |||||| |  |||| || ||| ||  || |||| || |
Db    296  TCATTGCCCATAGGCACCATTGTCATGGTGACGGGCTTCCGGGTGCCCGGGGCCATCA  355

Qy    361  CTGAGAGCAAGTGCTGCCCCGTCGGACGTTCTTCTCCATCCTCTCCTCTCCATTGCTG  420
           ||    || | || | || |  | ||||| ||| | ||| |  |  | |  |  | | |
Db    356  AGGAAAACAAGTGCTGCCCCTCCCTCCAGCTTTTTCATCGTCCTGTGGTCATCCTCCTAGCAG  415

Qy    421  AGGTTGCAGCTGCTGTGGTCGCCCTTGGTGTACA---CCACAATGGCTGAGCACTTCCTGA  477
           ||| |   ||| |   | ||   || || ||||       ||  ||| |   ||  ||  ||
Db    416  AGCTGATCTTACTCATCCTCTCTTTGTCTACATGGACAAGGTGAACGAAGAAGCCCAAGA  475

Qy    478  CGTTGCTGTGTAGTGCCTGCCATCAAGAAAGATATATGGTTCCCAGGAAGACTTCACTCAAG  537
           | |  ||  ||  ||  || || || |||| |  |  |  | | |||| || ||  ||| |
Db    476  AGGACCTGAAGGAAGGCCCTGCTGCTGCACCACGAGAACAACGTGGGCTGAAGAACG  535

Qy    538  TGTGGAACCACCACCATGAAAGGGCTCAAGTGCTGCTTCACCAACTATACGGATT  594
           |||||  | || || |||      |   ||  |||| ||||  || || ||  || |
Db    536  CCTGGAACATCATCCAGCTGACGGATGCGATGCGTGGTGCACTGACTACACCAGACT  592

RESULT  6
AF052492
LOCUS       AF052492     835 bp    mRNA         ROD       07-SEP-1998
DEFINITION  Mus musculus cell surface glycoprotein A15 mRNA, complete cds.
ACCESSION   AF052492
NID         g3549250
VERSION     AF052492.1   GI:3549250
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Hosokawa,Y., Senba,E. and Seto,M.
  TITLE     Molecular cloning of a cDNA encoding mouse A15, a member of
            transmembrane 4 superfamily and its preferential expression in
            brain neurons
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 835)
  AUTHORS   Hosokawa,Y., Senba,E. and Seto,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1998) Lab of Chemotherapy, Aichi Cancer Center
            Research Institute, Kanakoden 1-1, Chikusa-ku, Nagoya, Aichi
            464-8681, Japan
FEATURES             Location/Qualifiers
     source          1..835
                     /organism="Mus musculus"
                     /strain="C57BL/B6"
                     /db_xref="taxon:10090"
```

```
CDS                  /tissue_type="spleen"
                     6..740
                     /function="may be involved in cell proliferation and cell
                     motility"
                     /codon_start=1
                     /product="cell surface glycoprotein A15"
                     /protein_id="AAC34579.1"
                     /db_xref="PID:g3549251"
                     /db_xref="GI:3549251"
                     /translation="METKPVITCLKTLLIIYSFVWITGVILLAVGVWGKLTLGTYIS
                     LIAENSTNAPYVLIGTGTTIVVFGLFGCFATCRGSPWMLKLYAMFLSLVFLAELVAGI
                     SGFVFRHEIKDTFLRTYDAMQNYNGNDERSRAVDHVQPSLSCCGVQNYTNWSSSPYF
                     LDHGIPPSCCMETDCNPLDLHNLTVAATKVNQKGCYDLVTSFMETNMGIIAGVAFGI
                     AFSQIGMLLACCLSRFITANQYEMV"

BASE COUNT    191 a    205 c    214 g    225 t
ORIGIN

Query Match               5.7%;   Score 73.2;  DB 12;  Length 835;
Best Local Similarity    46.4%;   Pred. No. 1.2e-10;
Matches  332;  Conservative    0;  Mismatches  363;  Indels   21;  Gaps    2;

Qy    120  ACCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTT  179
           ||||                |   ||| || |||| || |  ||  ||  ||  ||  |||
Db     12  ACCAAACGTGATAAACCTGTCCAAAACCCTCCATCATCTACTCCTTCGTCTTCTGG  71

Qy    180  CTGTGTGGGTGCAGCCCTGTTGGCAGTGGGCATCTGGTGTCAATCGATGGGCCATCCTTT  239
           || |||||| ||||||  || ||| |||| | || |  | || ||| | | |||||||
Db     72  ATCACTGGGGTGATCCTGTTGGCCGTTGGCGTGTGGGGAAAGCTGACTTTGGGAACCTAT  131

Qy    240  CTGAAGATCTTCGGGCCACTGTCGTCCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTC  299
                                   || |   ||| | |||| |  ||  | |  |   |||  ||
Db    132  ------------ATCTCCCTGATTGCTGAGAACTCCACAAATGCTCCCTATGTG  173

Qy    300  CTCATCGCAGCCGGCGTTGTGGTCTTTGCTCCGGGTCCTGCTATGGTGCTAAG  359
           ||||| | | |||| |   ||||   || ||| ||| |  ||  | || | |||
Db    174  CTCATTGGAACGGGCACCACCATCGTGTTTTGGCCTCTTTGGATGCTTTGCACATGC  233

Qy    360  ACTGAGAGCAAGTGTGCCCCGTCGTGACGTTCTTCTTCCATCCTCTCCATCTTCATTGCT  419
           | ||     ||| ||  | || | | ||||| || ||| |  | |  |  | |  | ||  ||
Db    234  CGTGGTAGTCCATGGATGCTGAAACTGTATGCCATGTTCCTGTCCCTGGTGTTCCTGGCT  293

Qy    420  GAGGTTGCAGCTGCTGTGGTCGCCGTGGTTGGTGCATTCTGGATTTGTGTTTCGTCATGAGATCAAGGATCAAGGACACCTTCCTG  479
           || || |   || |  ||| | | ||| | | || | | ||| |  ||  |  |   |  |
Db    294  GAGCTTGTTGCTGGCCATTCTGGATTTGTGTTTCGTCATGAGATCAAGGATCAAGGACACCTTCCTG  353

Qy    480  TTGCTGTAGTGCCTGCCATCAAGAAAGATATGGTTCCCAGGAAGACTTCACTCAAGTG  539
           | | ||  ||  || || ||  ||| |  |  ||  | | || | ||  || ||| |
Db    354  AGGACTTACACGATGCCATGCAGAACTACAATGG---CAATGATGAGAGGAGCCCGGGCT  410

Qy    540  TGGAACACCACCATGAAAGGGCTCAAGTGCTGCTTCACCAACTATACGGATTTTGAG  599
                    ||      |||| |   |  |  |||| ||||  || || ||  || ||
Db    411  GTGGAACATGTGCAGCCGAGCCTGAGCTGTGGTGTGCAGAACTACACAAACTGGAGC  470

Qy    600  GACTCACCCTACTTCAAAGAGAACAGTGCCTTCCCCCATTCGTTGCAATGACAACGTC  659
                     ||||  |||| || || ||| || |  || |  | ||| || |  || |
Db    471  AGCAGCCCCTACTTCCTGGACCATGGCATCCCCCCAGTTGCTGCATGAACGAAACTGAC  530

Qy    660  ACCACACAGCCAATGAAACCTGACCAAGCAAAAGGCTCACGACCAAAAGTAGAGGGT  719
           ||  |  |  ||     |||  |  |  |   |||| | ||  || ||  | | ||
Db    531  TGCAATCCCCTGGATCTGCACAATCTGACTGTGGCCGCCACCAAAGTTAATCAGAAGGGC  590

Qy    720  TGCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGTGTGGCA  779
           ||    || | || ||| |  || | ||  || || || |  | ||||  |  ||  ||||
Db    591  TGTTATGATCTGGTGACCAGTTTTATGGAGACTAACATGGGGATCATTGCTGGAGTGGCA  650

Qy    780  GCTGAATTGGGGGCCCTGAGCTGCCATGATCTGTCCATGTATCTGTACTG  835
           |||||||  | ||  ||  |||| || |  | |  ||||  | | || ||
Db    651  TTTGGAATTGCGTTCTCCCAGTGATGTTGCATGCGTCTGCGGCTTGCTGTCTCGTCCCG  706

RESULT  7
HUMTAPA1
```

```
Db    674  GTGGTGAAGACCTTCCACGAGACGCTTGACTGCTGTGGCTCCAGCACCACTGACTGCTTTG  733

Qy    597  GAGGACTCA  605
                |||||
Db    734  ACCACCCTCA  742


RESULT  8
MUSTALLA
LOCUS       MUSTALLA       1720 bp    mRNA                    ROD         06-FEB-1999
DEFINITION  Mouse mRNA for PE31/TALLA, complete cds.
ACCESSION   D26483
NID         g685220
VERSION     D26483.1  GI:685220
KEYWORDS    PE31/TALLA; T-ALL associated antigen A15 couterpart; CCG-B7
            couterpart.
SOURCE      Mus musculus (strain:Balb/c) adult 8w whole brain cDNA to mRNA,
            clone_lib:library of M.Nagira clone:MPE31-1.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1720)
  AUTHORS   Ishikawa.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-1994) to the DDBJ/EMBL/GenBank databases. Izumi
            Ishikawa, Shionogi Institute for Medical Science; 2-5-1 Mishima,
            Settsu, Osaka 566, Japan (Tel:06-382-2612(ex.478), Fax:06-382-2598)
REFERENCE   2  (bases 1 to 1720)
  AUTHORS   Nagira,M., Ishikawa,I., Fujikawa,K., Takagi,S. and Yoshie,O.
  TITLE     Molecular Cloning and Expression of mouse PE31 (TALLA)
  JOURNAL   Unpublished (1994)
FEATURES             Location/Qualifiers
     source          1..1720
                     /organism="Mus musculus"
                     /strain="Balb/c"
                     /db_xref="taxon:10090"
                     /clone="MPE31-1"
                     /clone_lib="library of M.Nagira"
                     /dev_stage="adult 8w"
                     /tissue_type="whole brain"
     5'UTR           1..5
     CDS             6..740
                     /note="murine counterpart of A15 (T-ALL associated
                     antigen) and CCG-B7 which is expressed in the brain and is
                     containing CCG triplet repeats"
                     /codon_start=1
                     /product="murine PE31 (TALLA)"
                     /protein_id="BAA05493.1"
                     /db_xref="PID:d1006037"
                     /db_xref="PID:g685221"
                     /db_xref="GI:685221"
                     /translation="METKPVITCLRTLLIIYSFVFWITGVILLAVGVWGKLTLGTYIS
                     LIAENSTNAPVLIGTTIVVFGLFGCFATCRGSPWMLKLYAMFLSLVFLAELVAGI
                     SGFVPRHEIKDTFLRTYTDAMQNYNGNDERSRAVDHVQRSLSCCGVQNYTNWSSSPYF
                     LDHGIPPSCCMNETDCNPLDHNLTVAATKVNQKGCYDLVTSFMETNMGIIAGVAFGI
                     AFSQLIGMLLACCLSRFITANQYEMY"
     mat_peptide     6..737
                     /product="murine PE31/TALLA"
     3'UTR           740..1719
     polyA_signal    1698..1703
     polyA_site      1720
BASE COUNT      400 a        428 c       406 g       486 t
ORIGIN

Query Match              5.6%;       Score 71.6;  DB 12;  Length 1720;
Best Local Similarity   46.2%;       Pred. No. 3.8e-10;
Matches  331;  Conservative    0;  Mismatches  364;  Indels  21;  Gaps  2;

Qy   120  ACCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTT  179
               |||   ||| ||   | |||  ||   | |||    || | || | |   ||| ||
Db    12  ACCAAACCTGTGATAACCTGTCTCAAAAACCCTCCTCATCATCTACTCCTTCGTCTTCTGG  71
```

Left column:

```
LOCUS       HUMTAPA1       1496 bp    mRNA                    PRI         20-NOV-1992
DEFINITION  Human 26-kDa cell surface protein TAPA-1 mRNA, complete cds.
ACCESSION   M33680
NID         g338677
VERSION     M33680.1  GI:338677
KEYWORDS    26-kDa cell surface protein TAPA-1; target of antiproliferative
            antibody.
SOURCE      Human cell line OCI-LY8, cDNA to mRNA, clones 7-3 and 8-1.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1496).
  AUTHORS   Oren,R., Takahashi,S., Doss,C., Levy,R. and Levy,S.
  TITLE     TAPA-1, the target of an anti-proliferative antibody, defines a new
            family of transmembrane proteins
  JOURNAL   Mol. Cell. Biol. 10, 4007-4015 (1990)
  MEDLINE   90318365
COMMENT     Draft entry and computer readable sequence for [1] kindly submitted
            by S.Levy, 10-APR-1990, for release after publication.
FEATURES             Location/Qualifiers
     source          1..1496
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             239..949
                     /note="26-kDa cell surface protein TAPA-1"
                     /codon_start=1
                     /protein_id="AAA36663.1"
                     /db_xref="PID:g338678"
                     /db_xref="GI:338678"
                     /translation="MGVEGCTKCIKYLLFVFNFVFWLAGGVILGVALWLRHDPQTNL
                     LYLELGDKPAPNTFYVGIYILIAVGAVMMFVGFLGCYGAIQESQCLLGFFFTCLVILF
                     ACEVAAGIWGFVNKDDIAKDVKQFYDQALQQAVVDDDANNAKAVVKTFHETLDCCGSS
                     TLTALTTSVLKNNLCPSGSNIISNLFKEDCHQKIDDLFSGKLYLIGIAAITVAVIMIF
                     EMILSMVLCCGIRNSSVY"
     polyA_signal    1455..1460
BASE COUNT      257 a        504 c       413 g       322 t
ORIGIN

Query Match              5.6%;       Score 72.2;  DB 9;  Length 1496;
Best Local Similarity   49.9%;       Pred. No. 2.5e-10;
Matches  244;  Conservative    0;  Mismatches  233;  Indels  12;  Gaps  2;

Qy   129  TGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTCTGTGTGGT  188
               ||  |||| || || || |||| |||||| |||||||||||| || ||| |||||
Db    254  TGCACCAAGTGCATCAAGTACCTGCTCTTCGTCTTCAATTTCGTCTTCTGGCTGGCTGGA  313

Qy   189  GCAGCCCGTGTTGGCAGTGGGGCATCTGGGTGTCAATCGATGGG-------GAATCCTTTCTG  242
               || ||| ||| ||||||   ||| |||  ||| |      ||   ||||||| ||
Db    314  GGCGTGATCCTGGGTGTGGCCCTGTGGCTCCGCCATGACCCCCAGACCAACCTCCTG  373

Qy   243  AAGATCTTCGGGCACTGTCGTCCAGTGCCATGAGTCAGTTGTCAACGTGGCGTACTTCCTC  302
               || ||||  ||| || ||  ||   |||   |||| |     |   |||||| ||||
Db    374  TATCTGGAGCTGGGAGACAAGCCCGCGCCCAACACCTTCTATGTAGGCATCTACACCTC  433

Qy   303  ATCGCAGCCGCGGTTGTGGCTCTTGCCGTTCCGTTGGGCCTGTATGGTGCTAAGACT  362
               |||||| || ||||| || ||  | ||||||||||||||||||||
Db    434  ATCGCTGTGGGCGGCTGTCATGATGTTCGTTGGCTTCCTGGGCTGCTACGGGGCCATCCAG  493

Qy   363  GAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCCATCTTCATTGCTGAG  422
               ||  | || |  ||  || |||| ||| |||  |||   |||| ||||
Db    483  CTGGTAGTGCCTGCCATCAAGAA------AGATTATGGTCCCAGGAAGACTTCACTCAA  536

Qy   423  GTTGCAGCTGCTGTGGTGCGCCTTGGTGTGGGTGTACACCACAATGGCTGAGCACTTCCTGACGTTG  482
               |   |||  |||| || |||| | |||||  | ||    | || |||   |   ||
Db    614  TTCTATGACCAGGCCCTACACGGGATGTGGATGATGACGCCAACAACGCCAAGGCT  673

Qy   537  GTGTGGAACAACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCACTATACGGATTTT  596
               |||  |||| ||| |||| || ||  | | |||   | || ||   |
Db    12  ACCAAACCTGTGATAACCTGTCTCAAAAACCCTCCTCATCATCTACTCCTTCGTCTTCTGG  71
```

```
primer_bind     <1..14
CDS             1..711
                /codon_start=1
                /product="CD81"
                /protein_id="AAD11440.1"
                /db_xref="PID:g4206173"
                /db_xref="GI:4206173"
                /translation="MGVEGCTKCIKYLLFVFNFVFWLAGGVILGVALWLRHDPQTTNL
                LYLELGDKPAPNTFYVGIYILIAVGAVMMFVGFLGCYGAIQESOCLLGTFTCLVILF
                ACEVAAGIWGFVNKDQIAKDVKQFYDQALQQAVVDDDANNAKAVVKTHETLDCCGSS
                TLTALTTSVLKNNLCPSGSNIISNLFKEDCHQKIDDLFSGKLYLIGIAAIVVAVMIF
                EMILSMVLCCGIRNSSVY"
primer_bind     complement(693..>711)
BASE COUNT      136 a    212 c    200 g    163 t
ORIGIN

Query Match      5.5%;     Score 70.6;   DB 11;   Length 711;
Best Local Similarity  49.7%;   Pred. No. 6.4e-10;
Matches 243;   Conservative  0;   Mismatches 234;   Indels 12;   Gaps 2;
```

```
Qy  180  CTGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGGCATCCTTT  239
         | |||  | |||||||  || ||||||||| ||||||| | |||| ||||  |||
Db  72   ATCACTGGGGTGATCCTGTTGGCCGTTGGGAGTCTGGGAAAGCTGACTTTGGAACCTAT   131

Qy  240  CTGAAGATCTTCGGGCCACTGTCGTCCAGTGCCAGCAGTTGTTCAACGTGGCTACTTC     299
         |   ||| |||| |   ||| |||||  |||  ||||   |||  ||
Db  132  --------------ATCTCCCTGATTGCTGAGAACTCCACAAATGCTCCCTATGTG       173

Qy  300  CTCATCGCAGCCGGCCGTTGTGGTCTCTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAG  359
         |||| | | || | | |||||| ||  |  | ||| ||  |||| |  ||  | ||
Db  174  CTCATTGGAACCGGCCACCACCATCGTGGTTTTTGGCCTCTTTGGATGCTTTGCTACAATGC    233

Qy  360  ACTGAGAGCAAGTGCCCTCGTGACGTTGCCCTCGTCTTCTTCATCCTCCCTCATCTTCATTGCT  419
         ||  || | ||   | |      ||| |  ||| |
Db  234  CGTGGTAGTCCATGGATGCGTGAAACTGTGAAACTTCCTGTCCCTGTGTTCCTGGCCT      293

Qy  420  GAGGTTGCAGCTGCTGTGGTCGCCTTGGTGTACACCACAATGGCTGAGCACTTCCTGACG  479
         ||| || ||  | |||  || ||  |  |  |  ||| ||| | | |||| |   |||
Db  294  GAGCTTGTTGCTGGCATTTCTGGATTGTTGTTCTGCATGAGATCAAGGACACCTTCCTG    353

Qy  480  TTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTG  539
         | |  | ||  || |   | |  ||||  |||  | | |  ||
Db  354  AGGACTTACACGGATGCCGATGCCATGCAGTTACAAATGG---CAATGATGAGGAGCCGGGCT  410

Qy  540  TGGAACAACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAG  599
         ||||| || | ||| | ||||  ||| |||| | || |  | ||| ||| | ||
Db  411  GTGGACCATGTGCAGCGCGACGCTGAGCTGCGTGGCGTGTGCGAGAACTACACCACCTGGAGC  470

Qy  600  GACTCACCTACTTCAAAGAGAACAGAGTGCCTTTCCCCCATTCTCTTGCAATGACAACGTC  659
         |||||||| ||  ||||| ||
Db  471  AGCAGCCCCTACTTCCTGGACCATGGCATCCCCAGTTGCTGTGGCGCCCCAGTTAATCAGAAACTGAC  530

Qy  660  ACCACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGT  719
         ||  |  |  |    |   |  ||||  |  ||||  |  |  | ||||
Db  531  TGCAATCCCTGGATCTCGACCAGTTTTATGAGACTAACATGGGGATCATTGCTGGCATTGTCATCGTACTG  590

Qy  720  TGCTTCAATCAGCTTTGTTGTATGACAATCCGAACTAATGCAGTCAGTGCTGGTGTGTGGCA  779
         ||| |||||  || |  ||
Db  591  TGTTATGATCTGGTGACCAGTTTTATGAGACTAACATGGGGATCATTGCTGGCATTGTCATCGTACTGGCCA  650

Qy  780  GCTGGAATTGGGGGCCTCGAGCTGGCTGCTGCCAGCGATGATTGTGCCATGTATCTGTACTG  835
         |  ||||| | | | ||
Db  651  TTTGGAATTGCGTTCTCCCCAGTTGCATGGCATGGCATGCTGTTCTGCTTGCTGTCGTCTGTCCCCG  706
```

```
Qy  129  TGCTTCAGCTTCATTAAGACCATAGATGATCCTTCAATTGCTCATCTTCGTGTGTGGT  188
         |||  ||||  ||| |||||  | |||| |||| | |||||||||| ||| |||  ||
Db  16   TGCACCAAGTGCATCAAGTACCTGCTCTTCGTCTTCAATTTCGTCTTCTGGCTGGGA   75

Qy  189  GCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGG------GCATCCTTCTG  242
         | - ||| ||||| |  ||||  ||| ||| |||| |||||       |||||| |||
Db  76   GGGCGTGATCCTGGGTGTGGCCCTGTGGCTCCGGCATGACCCGCAGACCACCAACCTCTG  135

Qy  243  AAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCCTC  302
         |  || | || || |  |  || ||||  | |||| || ||  ||||| |||| |||||
Db  136  TATCTGGAGCTGGGAGACAAGCCCGCGCCACCTTCATGTAGGCATCATCTACATCCTC  195

Qy  303  ATCGCAGCCGCGTTGTGGCTCTTTGGTCTTCGGCTGCTGGGCTGCTATGGTGCTAAGACT  362
         ||  || | || |||  |  || ||| |  |||||  ||| | |  |  |  | ||||| |
Db  196  ATCGCTGTGGGGCGCGTCATGATGTTCGTCGGGGTTCCTGGGCTGCTACGGGGCCATCCAG  255

Qy  363  GAGAGCAAGTGCCCTCGTGACGTTCTTCTCAGCTGCATGGGCACTTCGAGCACTTCTGAG  422
         ||  |  |||  |  |  |||| |  | | |||  | | ||| |  | || |  | ||||
Db  256  GAATCCCAGTGCCTGTCGGGGGACGTTCTTCACCTGGCTGGCTGTCATCCTGTTCGGTGAG  315

Qy  423  GTTGCAGCTGCTGTGGTCGCCTTGGTGTCGCCTGGTACCACCACAATGGCTGAGCACGTTG  482
         ||| | ||| | | |||  ||| |  |   ||| | | |||||| |  | ||||  | |||
Db  316  GTGGCCGCCGCATCTGGGGCTTTGTGCAACAAGGACCAGATGCCGCCAAGGACGTGAAGCAG  375

Qy  483  CTGGTAGTGCCTGCCATCAAGAA------AGATTATGGTTCCCAGGAAGACTTCACTCAA  536
         | || |||  |||||||| | ||      ||| || || |  || |  ||| | |||
Db  376  TTCTATGACCAGGCCCTACAGCAGGCCGTGGTGATGACGCCAACACGCCAAGGCC  435

Qy  537  GTGTGGAACAACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTT  596
         ||  |  ||| ||| | |||| || | ||   | |||||  |  ||  |  | | ||
Db  436  GTGGTGAAGACCTTCCACGAGACGCTGGACTGCTGCGGCTCCAGCACACTGACTGCTTTG  495

Qy  597  GAGGACTCA  605
         || |||  |
Db  496  ACCACCTCA  504
```

```
RESULT 9
AF116600
LOCUS       AF116600      711 bp    mRNA            PRI       01-FEB-1999
DEFINITION  Pan troglodytes CD81 mRNA, complete cds.
ACCESSION   AF116600
NID         g4206172
VERSION     AF116600.1  GI:4206172
KEYWORDS    .
SOURCE      chimpanzee.
ORGANISM    Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 711)
AUTHORS     Levy,S. and Kuo,C.C.
TITLE       CD81 cDNA sequence derived from chimpanzee cells
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 711)
AUTHORS     Levy,S. and Kuo,C.C.
TITLE       Direct Submission
JOURNAL     Submitted (24-DEC-1998) Medicine/Oncology, Stanford, Stanford, CA
            94305, USA
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /cell_line="Janice"
                     /note="EBV-transformed cells; amplified using Homo sapiens
                     CD81-derived primers"

RESULT 10
AF116599
LOCUS       AF116599      711 bp    mRNA            PRI       01-FEB-1999
DEFINITION  Chlorocebus aethiops CD81 mRNA, complete cds.
ACCESSION   AF116599
NID         g4206170
VERSION     AF116599.1  GI:4206170
KEYWORDS    .
SOURCE      African green monkey.
ORGANISM    Chlorocebus aethiops
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;
            Chlorocebus.
REFERENCE   1  (bases 1 to 711)
AUTHORS     Levy,S. and Kuo,C.C.
```

TITLE   African green monkey CD81 cDNA
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 711)
AUTHORS   Levy,S. and Kuo,C.C.
TITLE   Direct Submission
JOURNAL   Submitted (24-DEC-1998) Medicine/Oncology, Stanford, Stanford, CA
             94305, USA
FEATURES   Location/Qualifiers
     source   1..711
                     /organism="Chlorocebus aethiops"
                     /db_xref="taxon:9534"
                     /cell_line="COS"
                     /note="amplified using Homo sapiens CD81-derived primers"
     primer_bind   <1..14
     CDS   1..711
                     /codon_start=1
                     /product="CD81"
                     /protein_id="AAD11439.1"
                     /db_xref="PID:g4206171"
                     /db_xref="GI:4206171"
                     /translation="MGVEGCTKCIKYLLFVFNFVFWLAGGVILGVALWLRHDPQTTNL
                     LYLELGDKPAPNTFYVGIYILIAVGAVMMFVGFLGCYGAIQESQCLLGTFFTCLVILF
                     ACEVAAGIWGFVNKDQIAKDVKQFYDQALQQAVVDDDANNAKAVVKTFHETLDCCGSS
                     TLAALTTSVLKNNLCPSGSNIISNLLKKDCHQKIDELFSGKLYLIGIAAIVVAVIMIF
                     EMILSMVLCCGIRNSSVY"
     primer_bind   complement(693..>711)
BASE COUNT   140 a   202 c   200 g   169 t
ORIGIN

Query Match   5.3%; Score 67.8; DB 11; Length 711;
Best Local Similarity 49.9%; Pred. No. 4.2e-09;
Matches 233; Conservative 0; Mismatches 222; Indels 12; Gaps 2;

Qy   128 GTGCTTCAGCTCATTAAGACCATGATGATCCTTCAATTTGCTCATCTTTCTGTGTGG 187
          ||| || ||| || ||| || ||| | || | | ||||| || || || |||
Db   15 GTGTACCAAGTGCATCAAGTACCTGCTCTTCGTCTTCAATTTCGTCTTCTGGCTGG 74

Qy   188 TGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGG------GCATCCTTTCT 241
                                                         | ||
Db   75 AGCCGTGATCCTGGTGTAGCCTTGTGCCTCCCCATGACCCCGCAGACTACCAACCTCCT 134

Qy   242 GAAGATCTTCGGGGCCACTGTCGTCGTGCCTCCAGTGCCATGCAGTTGTCAACGTGGCTACTTCCT 301
          ||| | | |||| || | || | | ||||||| ||| ||||| ||
Db   135 GTATCTCGAGCTGGGAGACAAGCCTGCACCCAATACCTTCTACGTAGGCATCTACATCCT 194

Qy   302 CATCGCAGCCGGCGTGTGGTCTTGCTGCGTCTTTGCTGGGCGTGCTATGGTGCTAAGAC 361
          ||| || || | | | ||| | | || || ||||| |||| |||| |
Db   195 CATTGCCGTGGGCGCTGTGATGATGTTCGTGGGCTTCCTGGGCTGCTACGGGGCCATCCA 254

Qy   362 TGAGAGCAAGTGCGCCCTCGTGACGTTCTTCTTCACCTGTCTGGTCATCCTGTTTGCCTGTGA 421
          |||| | ||| || | ||| || | || || | || | || || || | |||||||
Db   255 GGAATCCCAGTGCCTGCTGGGCACCTTCTTCACCTGTCTGGTCATCCTGTTTGCCTGTGA 314

Qy   422 GGTTGCAGCTGCGTGTGGTGCGCCTTGGTGTACACCAACAATGGCTGAGCACTTCCTGACGTT 481
          ||| |||| |||| | || | ||| |||||| | || ||| | | | ||||||||
Db   315 GCTTATCGTCGTTCTTCATCCTGCTGTGATTATCCTCCTTGCTGTGAGGTGACCTTGGCCAT 375

Qy   482 GCTGGTAGTGCCTGCCATCAAGAAAGATTATGTTCCAGGAAGACTTCACT------CA 535
          ||| | |||| |||| ||||| || ||||||||| ||||| |||||| |
Db   376 CCTGCTCTTGTATATGAACAGAAGCTGAATGAGTAGTGGC---TAAGGGTCTGACCGA 432

Qy   536 AGTGTGAACACCACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCA 582
          ||| || | || || | ||| | ||| | |||| || |||||||
Db   433 CAGCATCCACCGTTACACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCCATCCCAGTC 492

Qy   CGTGGTGAAGACCTTCCACGAGACGCCTTGACTGCTGTGCCTCCAGCA 481
           | ||| ||||| || ||||| | | ||| | |||| | ||||||
Db   435 AGGGCTCAAGTGCTGTGGCTTCACCAACTACACTACACGGATT 594

                                                              |||
Db   493 ATTTCTGCAGTGTGTGGTATAAATGGCACGAGTGATT 530

BASE COUNT   373 a   342 c   314 g   423 t
ORIGIN

Query Match   4.9%; Score 62.8; DB 9; Length 1452;
Best Local Similarity 51.2%; Pred. No. 1.3e-07;
Matches 173; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

Qy   257 ACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGCTACTTCCTCATCGCAGCCGGCGT 316
          | || |||| |||| || | |||||||| ||| | || ||| || || |||
Db   196 AGTGCTTCCATCAAACCTCCCCCCTCCTCAGCGGGCCAATGTGTTTGTCATCGTGGGGCTC 255

Qy   317 TGTGGTCTTTGCTCTTGGTTCCTGCTGCGTCTATGGTGCTAAGACTGAGAGCAAGTGC 376
          || | | ||| || || ||||||||| | ||| || || | || || |||| |
Db   256 TATTATCATGGTAGTTGCCGTTGCCTCTATCGGCTGCATCAAGGAAAACAAGTGTCT 315

Qy   377 CCTCGTGACGTTCTTCTTCACCTGTCTGGTCATCCTGTTTGCCTGTGA 421
          ||| |||| | || || | || | || || || | |||||||
Db   316 GCTTATCGTCGTTCTTCATCCTGCTGTGATTATCCTCCTTGCTGTGAGGTGACCTTGGCCAT 375

Qy   437 GGTCGCCTTGGTGTACACCAACAATGGCTGAGCACTTCCTGACGTTGCTGGTAGTGCCTGC 496
          || || ||| ||| | || ||| | | || || || | | | ||||| |||
Db   376 CCTGCTCTTGTATATGAACAGAAGCTGAATGAGTATGTGGC---TAAGGGTCTGACCGA 432

Qy   497 CATCAAGAAAGATTATGTTCCAGGAAGACTTCACTGGAACTGTGGAACACCACCATGAA 556
          ||| || || ||||| |||| ||| | || | || | | || | || || | |
Db   433 CAGCATCCACCGTTACACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCCATCCCAGTC 492

Qy   557 AGGGCTCAAGTGCTGTGGCTTCACCAACTACACTACACGGATT 594
          |||||| |||||||| ||||||||| |||| | | ||| ||
Db   493 ATTTCTGCAGTGTGTGGTATAAATGGCACGAGTGATT 530

VERSION   M60871.1 GI:180140
KEYWORDS   cell surface antigen; type III integral membrane protein.
SOURCE   Human promyelocytic tumor cell line HL60, cDNA to mRNA.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1452)
  AUTHORS   Amiot,M.
  TITLE   Identification and analysis of cDNA clones encoding CD53. A
             pan-leukocyte antigen related to membrane transport proteins
  JOURNAL   J. Immunol. 145 (12), 4322-4325 (1990)
  MEDLINE   91079522
FEATURES   Location/Qualifiers
     source   1..1452
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="promyelocytic tumor cell line HL60"
                     /map="Unassigned"
     mRNA   1..1452
                     /gene="CD53"
                     /note="G00-127-521"
     gene   1..1452
                     /gene="CD53"
     CDS   74..733
                     /gene="CD53"
                     /codon_start=1
                     /db_xref="GDB:G00-127-521"
                     /product="cell surface antigen"
                     /protein_id="AAA51951.1"
                     /db_xref="PID:g180141"
                     /db_xref="GI:180141"
                     /translation="MGMSSLKILLKYLVFFNLLFWICGCCILGFGIYLLIHNFGVLF
                     HNLPSLTLGNVFVIVGSIIMVAFLGCMGSIKENKCLLMSFPILLILILLAEVTLAIL
                     LFVYEQKLNEVVAKGLTDSIHRYHSDNSTKAAWDSIQSFLQCCGINGTSDWTSGPPAS
                     CPSDRKVEGCYAKARLWFHSNFLYIGIITICVCIEVLVGMSFALTLNCQIDKTSQTIG
                     L"

BASE COUNT   373 a   342 c   314 g   423 t
ORIGIN

RESULT 12
HUMCD53GLY
LOCUS   HUMCD53GLY   1480 bp   mRNA   PRI   03-MAR-1994
DEFINITION   Human CD53 glycoprotein mRNA, complete cds.

RESULT 11
HUMCD53
LOCUS   HUMCD53   1452 bp   mRNA   PRI   01-NOV-1994
DEFINITION   Human cell surface antigen (CD53) mRNA, complete cds.
ACCESSION   M60871
NID   g180140

ACCESSION   M37033
NID   g180142
VERSION   M37033.1 GI:180142
KEYWORDS   CD53 antigen.
SOURCE   Human lymphocyte LAK-cell, cDNA to mRNA, clone CD53-65.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1480)
AUTHORS   Angelisova.P., Vlcek,C., Stefanova,I., Lipoldova,M. and Horejsi,V.
TITLE   The human leucocyte surface antigen CD53 is a protein structurally
             similar to the CD37 and MRC OX-44 antigens
JOURNAL   Immunogenetics 32, 281-285 (1990)
MEDLINE   91055810
COMMENT   Draft entry and computer-readable sequence for [Immunogenetics
             (1990) In press] kindly submitted
             by V.Horejsi, 25-JUL-1990.
FEATURES   Location/Qualifiers
source   1..1480
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
CDS   94..753
             /codon_start=1
             /product="CD53 glycoprotein"
             /protein_id="AAA35663.1"
             /db_xref="PID:g180143"
             /db_xref="GI:180143"
             /translation="MGMSSLKLKLKYVLFFNLLFWICGCCILFGIYLLIHHNNFGVLF
             HNLPSLITLGNVFVIVGSIIMVVAFLGCMGSIKENKCLLMSFFILLLILLAEVTLAIL
             LFVYEQKLNEYVAKGLTDSIHRYHSDNSTKAAWDSIQSFLQCCGINGTSDWTSGPPAS
             CPSDRKVEGCYAKARLWFHSNFLYIGIITICVCVIEVLGMSFALTLNCQIDKTSQTIG
             L"
polyA_signal 1456..1461
BASE COUNT   378 a   353 c   318 g   431 t
ORIGIN

Query Match   4.9%;   Score 62.8; DB 9; Length 1480;
Best Local Similarity 51.2%;   Pred. No. 1.3e-07;
Matches 173; Conservative   0; Mismatches 162; Indels   3; Gaps   1;

Qy   257 ACTGTCGTCCAGTGCCATGCCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGGCGT 316
           | | | | || |||| | || |||||| ||| ||| || ||| || |||| | |||
Db   216 AGTGTCCTTCCATAACCTCCCCTCCCCCACGCGTGGGCAATGTGTTGTCATCGTGGGCTC 275

Qy   317 TGTGGTCCTTGCCTTCGTCGGTTCCTCGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGC 376
           | | ||||| | |||||||||| ||||||||||||| || || || |||||| ||||||||
Db   276 TATTATCATGGTAGTTGCCTTCCTGGGCTGCATGGGCTCTATCAAGGAGAAAACAAGTGTCT 335

Qy   377 CCTCGTGACGGTCTTCTCTTCATCCTCCTCATCTTCATTGCTGAGGTTGCAGCTGCTGT 436
           || | ||| | | ||||| | | | | || ||||||||||| ||| |||| ||
Db   336 GCTTATGTCGTTCTTCACCCTGCTGATTATCCTCCTTGCTGAGGTGACCTTGGCCAT 395

Qy   437 GGTCGCCTTGGTGTACACCACCAATGGCTGAGCACTTCCTGACGTTGCTGGTAGTGCCTGC 496
           |||| ||| ||||| | | || | | | ||||| |||| | || ||| ||||
Db   396 CCTGCTCTTTGTATATGAACAGAAGCTGAATGAGTATGTGGC---TAAGGGTCTGACCGA 452

Qy   497 CATCAAGAAAGATTATGGTTCCCAGGAGAGACTTCACTCAAGTGTGGAACACCCATGAA 556
           || | || | | | ||| ||| | ||||| |||| ||| || ||| | ||
Db   453 CAGCATCCACCGTTACCACTCAGACAATAGCACCAAGGCAGCGTGGGACTCCATCCAGTC 512

Qy   557 AGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATT 594
           || ||| || | | | |||| |||| | || ||||
Db   513 ATTTCTGCAGTGTGTGTGGTATAAAATGGCACGAGTGATT 550

RESULT 13
RNU19894
LOCUS   RNU19894   1303 bp   mRNA   ROD   25-MAR-1997
DEFINITION   Rattus norvegicus target of the antiproliferative antibody mRNA,
             complete cds.
ACCESSION   U19894
NID   g1142641
VERSION   U19894.1 GI:1142641
KEYWORDS   .
SOURCE   Norway rat.
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 1303)
AUTHORS   Geisert,E.E. Jr., Murphy,T.P., Irwin,M.H. and Larjava,H.
TITLE   A novel cell adhesion molecule, G-CAM, found on cultured rat glia
JOURNAL   Neurosci. Lett. 133 (2), 262-266 (1991)
MEDLINE   9227869
REFERENCE   2 (bases 1 to 1303)
AUTHORS   Irwin,M.H. and Geisert,E.E. Jr.
TITLE   The upregulation of a glial cell surface antigen at the astrocytic
             scar in the rat
JOURNAL   Neurosci. Lett. 154 (1-2), 57-60 (1993)
MEDLINE   93368802
REFERENCE   3 (bases 1 to 1303)
AUTHORS   Geisert,E.E. Jr., Yang,L. and Irwin,M.H.
TITLE   Astrocyte growth, reactivity, and the target of the
             antiproliferative antibody, TAPA
JOURNAL   J. Neurosci. 16 (17), 5478-5487 (1996)
MEDLINE   96346153
REFERENCE   4 (bases 1 to 1303)
AUTHORS   Geisert,E.E.
TITLE   Direct Submission
JOURNAL   Submitted (17-JAN-1995) Eldon E. Geisert Jr., Department of Anatomy
             and Neurobiology, University of Tennessee, Memphis, 855 Monroe
             Ave., Memphis, TN 38163, USA
FEATURES   Location/Qualifiers
source   1..1303
             /organism="Rattus norvegicus"
             /strain="Sprague-Dawley"
             /db_xref="taxon:10116"
             /clone="1.1, sp1, sp2"
             /clone_lib="Astrocyte cDNA lambda gt11 (Clontech No.
             RL1042B)"
             /sex="Male"
             /cell_type="astrocyte (Type 1 and Type 2)"
             /tissue_type="cerebral cortex"
5'UTR   1..222
CDS   223..933
             /note="TAPA"
             /codon_start=1
             /product="target of the antiproliferative antibody"
             /protein_id="AAC53103.1"
             /db_xref="PID:g1142642"
             /db_xref="GI:1142642"
             /translation="MGVEGCTKCIKYLLFVFNVFWLAGGVILGVALWLRHDPQTTTL
             LYLELGDKPAPSTFYVGIYILIAVGAVMMVFGFLGCYGAIQESQCLLGTFFTCLVILF
             ACEVAAGIWGFVNKDQIAKDVKQFYDQALQQAVMDDDANNAKAVVKTFHETLNCCGSN
             TLTLTTAVLRNSLCPSSSNSFTQLLKEDCHQKIDELFSGKLYLIGIAAIVVAVIMIF
             EMILSMVLCCGIRNSSVY"
3'UTR   934..1303
BASE COUNT   257 a   375 c   347 g   324 t
ORIGIN

Query Match   4.8%;   Score 62.2; DB 12; Length 1303;
Best Local Similarity 53.5%;   Pred. No. 1.9e-07;
Matches 177; Conservative   0; Mismatches 148; Indels   6; Gaps   2;

Qy   129 TGCTTCAGTTCATTAAGACCATGATGAGATCCCTTCAATTTGCTCATCTTTCTGTGTGGT 188
           ||| |||||| ||| || |||||| | || |||||||| ||||||||| || |||||||
Db   238 TGCACCAAATGCATCAAATACCTGCTCTTCGTCTTCAATTTCGTCTTCTGGCTGGCTGGA 297

Qy   189 GCAGCCCGTTGGGCAGTGGCATCTGGGTGTCAATCGATGGGGCATCCTTCTGCAGAGATC 248
           || || | |||| ||||| || | |||| || || | |||| || | | | |||
Db   298 GGTGTGATCCTAGGTGTAGCTCTGTGGTTGCGCCATGATCCACAGACCACCACCTTGCTC 357

Qy   249 TTCGGCCACTG----TCGTCCAGTGCCATGCAGTTTGTCAAGTG-GGCTACTTCCTC 302
           || |||| |||| ||| || | | ||| | || | |||||| || | |
Db   358 TACCTGGAACTGGGAGACAGAACCAGCACCTAGCACCTTCATGTGGGCATTCTACATTCTC 417

```
Matches 324;  Conservative  0;  Mismatches  371;  Indels  21;  Gaps  2;

Qy   120 ACCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTT 179
         |||| |||| |||| || |||  |||| |||  ||| ||| | ||| ||||| |||
Db    31 ACCAAACCTGTGATAACCTGCTCTCAAAACCCTCGCCATCATCTACTCCTCGTCTTCTGG 90

Qy   180 CTGTTGGTCAGCCCTGTTGGCAGTGGGCATCTGGCGTCAATCGATGGGGCATCCTTT 239
         |||  |||| |||||||| || ||| |||| ||| || ||| ||
Db    91 ATCACTGGGGTGATCCTGCCTGGCGTTGGGAGTCTGGGCAAACTTACTCTGGGCACCTAT 150

Qy   240 CTGAAGATCTTCGGGGACCACTGTCCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTC 299
         |||  ||| || |||||  ||| ||| ||||||
Db   151 --------ATCTCCCTATTGCCAGAACTCCACAAAATGCTCCCCTATGTG 192

Qy   300 CTCATCGCAGCCGGCGGTTGTGGTCTTGTCTTTGCTCTTGGTCTTCCTGGGCTGCATGGTGCTAAG 359
         ||||| || ||| | | |  ||| || || ||| ||  || ||| |||| |||| || |||||
Db   193 CTCATCGGAACTGGCACCACTATTGTGTCTCTTGGCCTGTTTGGCCTTCATCTTCACATGT 252

Qy   360 ACTGAGAGCAAGTGTCCCCTCGTGACGTTCTTCATCCTCCTCATCTTCATTGCT 419
         ||| || ||| || |  ||  |  || ||| ||||  |||| || ||
Db   253 CGTGGTAGCCCATGGATGCTGAAACTGTATGCCATGTTTCTGTCCCTGGTGTTCCTGGCT 312

Qy   420 GAGGTTGCAGCTGCTGGTGCGCCCTGGTGTACACCACCATGGCTGAGCACTTCCTGACG 479
         ||| ||||| ||| || ||
Db   313 GAGCTCGTAGCTGCTGGCATTCAGGGTTTGTGTTTCGTCATGAGATCAAGGACCACCTTCCTG 372

Qy   480 TTGCTGTGAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTG 539
         || |  || | | | | | |  | ||| |
Db   373 AGGACTTACACGGACGCTATGCAGACTTACAATGG---CAATGATGAGAGAGCGGGGCA 429

Qy   540 TGGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAG 599
         || ||| | || |  || ||| |||
Db   430 GTGGACCCATGTGCAGCGCAGCCTGAGCTGCTGGTGGTGCAGAACTACACCAACTGGAGC 489

Qy   600 GACTCACCTACTTCAAAGAGACAACAGTGCCCTTCCCCATTCGTGTGCAATGACAACGTC 659
         | |||||||   |  |||||   ||  | ||| ||
Db   490 ACCAGCCCTACTTCCTGGACATGGCATCCCCCAGCTGCTGCTGCATGAGCAACGAAACTGAT 549

Qy   660 ACCAACACAGCCAATGAAACCTGCACCAAGCAAAGGCTCACGACCACCAAAGTAACCAGAGGGT 719
         ||  || |  |||  || || |||||||| |||
Db   550 TGTAATCCCCAGGATCTACACAATCTACTGTGCCGCCGCCACCAAAGTAACCAGAGGGT 609

Qy   720 TGCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTCGGGTGGTGTGGCA 779
         || || |  |  || |||  |||    |
Db   610 TGTTATGATCGGTAACTAGTTTCATGGAGACTAACATGGGAATCATCGCTGGAGTGGCG 669

Qy   780 GCTGGAATTGGGGCCTCAGCTGGCCTGCCATGATTGTGTCCATGATTATCTGTACTG 835
         ||  |||| ||| || ||||
Db   670 TTTGGAATCGGATTCTCCCAGTTAATTGGCATGCTGCTGGCCTGCTGCCTGTCTCGTTTCCG 725

RESULT 14
HUMA15
LOCUS       HUMA15        1743 bp     mRNA              PRI       02-FEB-1999
DEFINITION  Homo sapiens mRNA for cell surface glycoprotein, complete cds.
ACCESSION   D10653
NID         g285900
VERSION     D10653.1    GI:285900
KEYWORDS    ME491/CD63 superfamily; cell surface glycoprotein.
SOURCE      Homo sapiens immature T cell lymphocyte cell_line:HPB-ALL cDNA to
            mRNA, clone_lib:lambda gt10 clone:A15.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1743)
AUTHORS     Emi,N.
TITLE       Direct Submission
JOURNAL     Submitted (29-FEB-1992) to the DDBJ/EMBL/GenBank databases.
            Nobuhiko Emi, Nagoya University School of Medicine, First Dept. of
            Internal Medicine; 65 Tsurumai, Showa-ku, Nagoya, Aichi 466, Japan
            (Tel:052-741-2111, Fax:052-741-1612)
REFERENCE   2  (bases 1 to 1743)
AUTHORS     Emi,N., Kitaori,K., Seto,M., Ueda,R., Saito,H. and Takahashi,T.
TITLE       Isolation of a novel cDNA clone showing marked similarity to
            ME491/CD63 superfamily
JOURNAL     Immunogenetics 37 (3), 193-198 (1993)
MEDLINE     93131291
COMMENT     On Jun 11, 1993 this sequence version replaced gi:219404.
FEATURES             Location/Qualifiers
     source          1..1743
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HPB-ALL"
                     /product="cell surface glycoprotein"
                     /protein_id="BAA01501.1"
                     /db_xref="PID:d1001976"
                     /db_xref="PID:q285901"
                     /cell_type="lymphocyte"
                     /clone="A15"
                     /db_xref="GI:285901"
                     /clone_lib="lambda gt10"
                     /tissue_type="immature T cell"
     gene            25..759
                     /gene="A15"
     CDS             25..759
                     /gene="A15"
                     /codon_start=1
                     /product="cell surface glycoprotein"
                     /translation="METKPVITCLKTLLIIYSFVFWITGVILLAVGVWGKLILGTYIS
                     LIAENSTNAPYVLIGTGTTIVVFGLFGCFATCRGSPWMLKIYAMFLSLVFLAELVAGI
                     SGFVFRHEIKDIFLRYYTDAMQTYNGNDERSRAVDHVQRSLSCCGVQNYTNWSTSPYF
                     LEHGIPPSCCMNETDCNPQDLHNLTVAATKVNQKGCYDLVTSFMETNMGIIAGVAFGI
                     AFSQLIGMLLACCLSRFITANGYEMV"
     polyA_signal    1722..1727
     polyA_site      1743
BASE COUNT      432 a    424 c    384 g    503 t
ORIGIN

RESULT 15
HUMRNAB
LOCUS       HUMRNAB       1792 bp     mRNA              PRI       22-JUN-1993
DEFINITION  Human (clone CCG-B7) mRNA sequence.
ACCESSION   L10373
NID         g307287
VERSION     L10373.1    GI:307287
KEYWORDS    .
SOURCE      Homo sapiens (library: Stratagene lambda ZAP II) female 85 years
            frontal Cortex cDNA to mRNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1792)
AUTHORS     Li,S.-H., McInnis,M., Margolis,R., Antonarakis,S. and Ross,C.
TITLE       Novel triplet repeat containing genes in human brain: Cloning,
            expression, and length polymorphisms
JOURNAL     Genomics 16, 572-579 (1993)
MEDLINE     93315145
FEATURES             Location/Qualifiers
     source          1..1792
                     /organism="Homo sapiens"
```

Qy   303 ATCGCAGCCGGCGGTTGTGGTCTTGTCTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGACT 362
         ||| ||| || || || | ||| || |||||| | | |||||||||| |
Db   418 ATTGCTGTGGGAGCTGTGATGATGTTTGTGAGGCTTCGTGGGTGCTCATGGGGCCATCCAG 477

Qy   363 GAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTCCTGAG 422
         ||| |||| || | ||| | |||||||||| ||| || | |||| |||||
Db   478 GAGTCCCAGTGCGTTGCTGCGACGTTCTTCACTGCGTTCTCTGTTTGCCTGTGAG 537

Qy   423 GTTGCAGCTGCTCTGTGTCGCCTTGGTGTACA 453
         ||| ||| || ||| | ||
Db   538 GTAGCTGCGGGGCATCGGGGCTTTGTAAACA 568

```
                    /db_xref="taxon:9606"
                    /dev_stage="85 years"
                    /sex="female"
                    /tissue_type="frontal Cortex"
                    /tissue_lib="stratagene lambda ZAP II"
BASE COUNT     434 a     447 c     399 g     512 t
ORIGIN

Query Match          4.7%;    Score 60.4;  DB 9;   Length 1792;
Best Local Similarity 45.3%;  Pred. No. 6.8e-07;
Matches 324; Conservative   0; Mismatches 371; Indels   21;  Gaps    2;

Qy  120 ACCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTTCATCTTT 179
            ||||  | ||   |  | || || ||| | ||| ||| ||  |  || |||
Db   75 ACCAAACCTGTGATAACCTGTCTCAAAACCCTCCTCAATCATCTACTCCTTCGTCTTCTGG 134

Qy  180 CTGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTT 239
          | |||| ||||||| ||  || || ||| || |||  | || ||||  |||  |||||
Db  135 ATCACTGGGGTGATCCTGCTGGCTGTTGGAGTCTGGGCAAACTTACTCTCGGGCACCTAT 194

Qy  240 CTGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTTGTCAACGTGGCTACTTC 299
            ||  |  |  |||   ||| ||| || |  | |  ||| |  |||||
Db  195 --------------ATCTCCCTTATTGCCGAGAACTCCACAAATGCTCCCTATGTG 236

Qy  300 CTCATCGCAGCCGGCCGTGTGGTCTTTGCCTCTTTGGTTTCCTGGGCTGCTATGGTGCTAAG 359
          ||||||| || |  | | |  |    || ||| || ||||  || |||||
Db  237 CTCATCGGAACTGGCACCACTATTGTTGCTTTGGCCTGTTTGGATGCTTGCTACACATGT 296

Qy  360 ACTGAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCT 419
          ||   ||||    || ||   |  ||| ||||   | |  | ||||  ||  |
Db  297 CGTGTGTAGCCCATGGATGCTGAAACTGTATGCCATGTTTCGTCCCTGGTGTTCCTGGCT 356

Qy  420 GAGGTTGCAGCTCGCTGTGGTCGCCTGGTACACCACAATGGCTGAGCACTTCCTGACG 479
          || ||| ||||||  |  || || |    ||| |||| |||  ||||||  |||||||
Db  357 GAGCTCGTAGCTGGCCATTTCAGGGTTTGTTTCGTCATGAGATCAAGGACACACCTTCCTG 416

Qy  480 TTGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTG 539
          | |  ||||| ||   |  |||| |  | |     ||||||   |  |    | |
Db  417 AGGACTTACACGGACGCTATGCAGACTTACAATGG----CAATGATGAGAGGAGCCGGGCA 473

Qy  540 TGGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAG 599
          |||  || |||| |||| |  || || |   || |    | |   |   || || |
Db  474 GTGGACCATGTGCAGCGCAGCCTGAGCTGTGGTGCTGGTGGTGTGCAGAACTACACCACTGGAGC 533

Qy  600 GACTCACCCTACTTCAAAGAGACAACAGTGCCTTTCCCCATTCGTGTTGCAATGACAACGTC 659
          |   || ||| | |||||  |      || |    | ||| |  |  || || ||  ||||
Db  534 ACCAGCCCCTACTTCCTGGAGCATGGCATCCCCCCCAGCTGCTGCATGAACGAAACTGAT 593

Qy  660 ACCACACAGCCAATGAAACCTGACCAAGCAAAGGCTCACGACCAAAAAGTAGAGGGT 719
          || |  |    | |||  |  | || |  ||| |  |   |     |    ||||||
Db  594 TGTAATCCCAGGATCTACACAATCTGACTGGCCGCCACCAAAGTTAACCAGAAGGGT 653

Qy  720 TGCTTCAATCAGCTTTTGTATGACACATCCGAACTAATGCAGTCACCGTGGTGTGTGGCA 779
          || |  |  ||||  |  ||  | |  |||||  |  |   | || |||| |||||
Db  654 TGTTATGATCTGGTAACTAGTTTCATGGAGACTAACATGGAATCATCCGCTGGGAGTGGCG 713

Qy  780 GCTGGAATTGGGGCCTCGAGCCTGCCTGGCTGCCATGATTGTGTCCATGTATCGTACTG 835
          ||| ||| |    ||  |  ||| | ||  |     | || |  |  | |  | || ||
Db  714 TTTGGAATCGCATTCTCCCCAGTTAATTGGCATGGCATGCGTCTGGCCTGCTGTCGTCCCG 769

Search completed: September 28, 1999, 12:21:38
Job time: 4978 sec
```

This Page Blank (uspto)

```
                                                                              44    44.2   3.4   1158  1  Q37067      SJ23-like protein
                                                                              45    40     3.1   1694  1  V33190      Secreted protein B
```

OM nucleic - nucleic search, using sw model

ALIGNMENTS

Copyright   GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: September 28, 1999, 12:26:53 ; Search time 289.74 Seconds
                                          (without alignments)
                                          1113.059 Million cell updates/sec

Title: US-09-030-606-111
Perfect score: 1289
Sequence: 1 AGCCAGGCGTCCCTCGCCT.........GTTAAAAAAAAAAAAAAAA 1289

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database :   N_Geneseq_36:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

SUMMARIES

```
                    %
Result          Query
No.    Score   Match  Length DB  ID       Description
-----  -----   -----  ------ --  ------   --------------------------
   1   1289    100.0   1289   1  V58587   Prostate tumour sp
   2   1289    100.0   1289   1  V61202   Full length cDNA s
   3   1277.6   99.1   1376   1  V34157   Human secreted pro
   4   1274.8   98.9   1605   1  V30916   Human secreted pro
   5   1273.6   98.8   1324   1  V34225   Human secreted pro
   6   1262     97.9   1288   1  V48113   Nucleotide sequenc
   7    928     72.0    933   1  V54014   Nucleotide sequenc
   8    602.4   46.7    801   1  V58493   3' fragment of pro
c  9    602.4   46.7    801   1  V61150   5' cDNA sequence o
c 10    570.8   44.3    740   1  V58494   5' fragment of pro
  11    570.8   44.3    740   1  V61151   5' cDNA sequence o
  12    560.8   43.5    729   1  V58490   5' fragment of pro
  13    560.8   43.5    729   1  V61147   5' cDNA sequence o
  14    528.4   41.0    751   1  V58489   3' fragment of pro
c 15    528.4   41.0    751   1  V61146   3' cDNA sequence o
c 16    493.4   38.3   1494   1  V34226   Human secreted pro
  17    411.4   31.9    427   1  X41224   Human secreted pro
  18    394.8   30.6    415   1  V86414   EST clone AR415. N
  19    376.2   29.2    433   1  X40584   Human secreted pro
  20    307.2   23.8    378   1  X51959   Human secreted pro
  21    296     23.0    326   1  X40605   Human secreted pro
  22    282.8   21.9    297   1  V15588   Human HPK-1A C4.8
  23    248.8   19.3   1890   1  V80623   Kidney injury asso
  24    115      8.9    127   1  T23814   Human gene signatu
  25     83.4    6.5    977   1  X16296   Human CD53-like tr
  26     81.4    6.3   1345   1  V48117   Nucleotide sequenc
  27     75.4    5.8   1151   1  V61010   Human integral mem
  28     72.2    5.6    936   1  V48120   Nucleotide sequenc
  29     63.8    4.9   1849   1  V48111   Nucleotide sequenc
  30     62.8    4.9   1452   1  Q21187   Encodes CD53 haema
  31     62.8    4.9   1452   1  T14726   Human CD53 antigen
  32     62.8    4.9   1452   1  V81220   Human CD53 antigen
  33     62.8    4.9   1452   1  V63463   Human CD53 antigen
  34     57.6    4.5    734   1  V48118   Nucleotide sequenc
  35     56      4.3   1120   1  Q29182   DNA encoding metas
  36     55.4    4.3   1624   1  T40021   Human metastasis t
  37     55      4.3    430   1  V88180   EST clone DW181. N
  38     48.6    3.8   1104   1  Q22438   CO-029 tumour asso
  39     48.4    3.8    990   1  X37471   Human secreted pro
  40     47.6    3.7   1722   1  T84982   Human membrane ant
  41     47.6    3.7   1662   1  V48116   Nucleotide sequenc
  42     47.6    3.7   1722   1  X21973   HP00966 coding seq
  43     46      3.6    428   1  X41047   Human secreted pro
```

RESULT 1

V58587
ID   V58587 standard; cDNA; 1289 BP.
AC   V58587;
DT   08-DEC-1998 (first entry)
DE   Prostate tumour specific gene clone N1-1862.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   123..848
FT   /*tag= a
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1999; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
DR   P-PSDB; W69386.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 85; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 1289 BP; 292 A; 335 C; 328 G; 334 T;

Query Match   100.0%; Score 1289; DB 1; Length 1289;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 1289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 AGCCAGGCGTCCCTCGCCTGCCCACTCAGTGGCAACACCCGGAGCTGTTTGTCCTTT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AGCCAGGCGTCCCTCGCCTGCCCACTCAGTGGCAACACCCGGAGCTGTTTGTCCTTT 60

Qy   61 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 120

Qy  121 CCATGCAGTGCTTCAGCTTCATTAAGAACCATGATGATCCTCTTCAATTGCTCATCTTC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CCATGCAGTGCTTCAGCTTCATTAAGAACCATGATGATCCTCTTCAATTGCTCATCTTC 180

Qy  181 TGTGTGGTGCAGCCCTGTTGGCAGTGGCAGTCTCGGGTGTCAATCGATGGGGCATCCTTC 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TGTGTGGTGCAGCCCTGTTGGCAGTGGCAGTCTCGGGTGTCAATCGATGGGGCATCCTTC 240

Qy  241 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC 300

Qy  301 TCATCGCAGCCGGCGGTGTGGTCTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TCATCGCAGCCGGCGGTGTGGTCTTTGCTCTTGGTTCCTGGGCTGCTATGGTGCTAAGA 360

Qy  361 CTGAGAGCAAGTGCCCCTCGTGACGTTCTTCTTCATCCTCCTCCATCTTCATTGCTG 420
```

```
PD   27-AUG-1998.
PF   25-FEB-1998.  U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609986/51.
DR   P-PSDB; W71870.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 80; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 1289 BP;     292 A;     335 C;     328 G;     334 T;

Query Match             100.0%;  Score 1289;  DB 1;  Length 1289;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1289;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1  AGGCCAGGCGTCCCTCCTCCTGCCCACTCCAGTGCAACACCCGGAGCTGTTTGTCCTTT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  AGGCCAGGCGTCCCTCCTCCTGCCCACTCCAGTGCAACACCCGGAGCTGTTTGTCCTTT   60

Qy     61  GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA  120

Qy    121  CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTTCAATTGCTCATCTTTC    180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTTCAATTGCTCATCTTTC    180

Qy    181  TGTGTGGTGCAGCCCTGTTGGCAGTGGCAGTGGGCATCGATGGGGCATCCTTTC     240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TGTGTGGTGCAGCCCTGTTGGCAGTGGCAGTGGGCATCGATGGGGCATCCTTTC     240

Qy    241  TGAAGATCTTCGGGCCACTGTCGTCGTCCAGTGTTGTCAACGTGGGCTACTTCC      300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TGAAGATCTTCGGGCCACTGTCGTCGTCCAGTGTTGTCAACGTGGGCTACTTCC      300

Qy    301  TCATCGAGCCGCGTGTGGTCTTTGCTCTGTGGTTTCCTGGCGTGCTGTATGGTGCTAAGA  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  TCATCGAGCCGCGTGTGGTCTTTGCTCTGTGGTTTCCTGGCGTGCTGTATGGTGCTAAGA  360

Qy    361  CTGAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTTCCTCCTCCTCCTCCATCTTCATTGCTG  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  CTGAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTTCCTCCTCCTCCTCCATCTTCATTGCTG  420

Qy    421  AGGTTGCAGCTGCTGTGTCGCCGTTGGTCGCCTTGGTGTACACCACCAATGGCTGCACTTCCTGACGT  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AGGTTGCAGCTGCTGTGTCGCCGTTGGTCGCCTTGGTGTACACCACCAATGGCTGCACTTCCTGACGT  480

Qy    481  TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT  540

Qy    541  GGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGG  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGG  600

Qy    601  ACTCACCCTACTTCAAAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  ACTCACCCTACTTCAAAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA   660

Qy    661  CCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTT  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTT  720
```

Left column continuation (rotated):

```
Db        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          CTGAGAGCAAGTGTGCCCCTCGTGACGTTCTTCTTCCTCCTCCTCCTCCATCTTCATTGCTG   420
Qy    421  AGGTTGCAGCTGCTGTGTCGCCGTTGGTCGCCTTGGTGTACACCACCAATGGCTGCACTTCCTGACGT   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AGGTTGCAGCTGCTGTGTCGCCGTTGGTCGCCTTGGTGTACACCACCAATGGCTGCACTTCCTGACGT   480
Qy    481  TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT   540
Qy    541  GGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGG   600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGG   600
Qy    601  ACTCACCCTACTTCAAAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA   660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  ACTCACCCTACTTCAAAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA   660
Qy    661  CCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTT   720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGGTT   720
Qy    721  GCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCAGTGGGGCACCGTG    780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  GCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCAGTGGGGCACCGTG    780
Qy    781  CTGGAATTGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGTATCTGTACTGCAATC   840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  CTGGAATTGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGTATCTGTACTGCAATC   840
Qy    841  TACAATAAGTCCACTTCGTCCCTCTGCCCTCGCCACTACTGCCCACTACTGGGAACTGTGAAGAGGC  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  TACAATAAGTCCACTTCGTCCCTCTGCCCTCGCCACTACTGCCCACTACTGGGAACTGTGAAGAGGC  900
Qy    901  ACCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA   960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  ACCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA   960
Qy    961  GAATGGACCTGCCCTTCTGCTCCAGTCCAGACTTGGGCTAGATAGGGACCCACTCCTTTAGCG  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  GAATGGACCTGCCCTTCTGCTCCAGTCCAGACTTGGGCTAGATAGGGACCCACTCCTTTAGCG  1020
Qy   1021  ATGCCTGACTTTCCTTCCTTCTGTTGGTGGGTGGGATGGGTGGGGGTGGGGGCCTCTAAG  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  ATGCCTGACTTTCCTTCCTTCTGTTGGTGGGTGGGATGGGTGGGGGTGGGGGCCTCTAAG  1080
Qy   1081  GTAGCCAGTTCTGTTGCCCATTCCCCCAGTCTATTAAACCCTTGATATGCCCCCTAGGCC  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  GTAGCCAGTTCTGTTGCCCATTCCCCCAGTCTATTAAACCCTTGATATGCCCCCTAGGCC  1140
Qy   1141  TAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTATAGCCTGGGCAT  1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  TAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTATAGCCTGGGCAT  1200
Qy   1201  AAGTGAAATCAGCAGAGCCTCTCGGTGATGTGTAGAAGGCACTTCAAAATGCATAAACC  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  AAGTGAAATCAGCAGAGCCTCTCGGTGATGTGTAGAAGGCACTTCAAAATGCATAAACC  1260
Qy   1261  TGTTACAATGTTAAAAAAAAAAAAAAAAA 1289
          |||||||||||||||||||||
Db   1261  TGTTACAATGTTAAAAAAAAAAAAAAAAA 1289
```

```
RESULT   2
V61202
ID   V61202 standard; cDNA; 1289 BP.
AC   V61202;
DT   06-JAN-1999  (first entry)
DE   Full length cDNA sequence of prostate tumour clone N1-1862.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO983709 3-A2.
```

```
Qy   721 GCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCACCGTGGGTGGTGTGGCAG  780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCACCGTGGGTGGTGTGGCAG  780

Qy   781 CTGGAATTGGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGTATCTGTACTGCAATC  840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CTGGAATTGGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGTATCTGTACTGCAATC  840

Qy   841 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGAACTGTGAAGAGGC  900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGAACTGTGAAGAGGC  900

Qy   901 ACCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA  960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 ACCCTGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA  960

Qy   961 GAATGGACCTGCCCTTTCTGCTCCCAGACTTGGGGCTAGATAGGGACCACTCCTTTTAGCG 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GAATGGACCTGCCCTTTCTGCTCCCAGACTTGGGGCTAGATAGGGACCACTCCTTTTAGCG 1020

Qy  1021 ATGCCTGACTTTCCTTCCATTGGTGGGTGGATGGGTGGGGGCATTCCAGAGCCTCTAAG  1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ATGCCTGACTTTCCTTCCATTGGTGGGTGGATGGGTGGGGGCATTCCAGAGCCTCTAAG  1080

Qy  1081 GTAGCCAGTTCTGTTGCCCATTCCCCAGTCTCTATTAAACCCTTGATATGCCCCCTAGGCC 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GTAGCCAGTTCTGTTGCCCATTCCCCAGTCTCTATTAAACCCTTGATATGCCCCCTAGGCC 1140

Qy  1141 TAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT  1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 TAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT  1200

Qy  1201 AAGTGAAATCAGCAGAGCCTCTGGGTGATGTGTAGAAGGCACTTCAAAATGCATAAACC  1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AAGTGAAATCAGCAGAGCCTCTGGGTGATGTGTAGAAGGCACTTCAAAATGCATAAACC  1260

Qy  1261 TGTTACAATGTTAAAAAAAAAAAAAAAA 1289
         ||||||||||||||||||||||||||||
Db  1261 TGTTACAATGTTAAAAAAAAAAAAAAAA 1289
```

```
RESULT   3
V34157
ID   V34157 standard; DNA; 1376 BP.
AC   V34157;
DT   28-JAN-1999  (first entry)
DE   Human secreted protein gene 4 clone HKCSR70.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9839446-A2.
PD   11-SEP-1998.
PF   06-MAR-1998;  U04492.
PR   07-MAR-1997; US-038621.
PR   07-MAR-1997; US-040161.
PR   07-MAR-1997; US-040162.
PR   07-MAR-1997; US-040163.
PR   07-MAR-1997; US-040333.
PR   07-MAR-1997; US-040334.
PR   07-MAR-1997; US-040336.
PR   07-MAR-1997; US-040626.
PR   11-APR-1997; US-043311.
PR   11-APR-1997; US-043312.
PR   11-APR-1997; US-043313.
PR   11-APR-1997; US-043314.
PR   11-APR-1997; US-043315.
PR   11-APR-1997; US-043568.
PR   11-APR-1997; US-043569.
PR   11-APR-1997; US-043576.
PR   11-APR-1997; US-043578.
PR   11-APR-1997; US-043580.
PR   11-APR-1997; US-043369.
PR   11-APR-1997; US-043670.
PR   11-APR-1997; US-043671.
PR   11-APR-1997; US-043672.
PR   11-APR-1997; US-043674.
PR   23-MAY-1997; US-047492.
PR   23-MAY-1997; US-047500.
PR   23-MAY-1997; US-047501.
PR   23-MAY-1997; US-047502.
PR   23-MAY-1997; US-047503.
PR   23-MAY-1997; US-047581.
PR   23-MAY-1997; US-047582.
PR   23-MAY-1997; US-047583.
PR   23-MAY-1997; US-047584.
PR   23-MAY-1997; US-047585.
PR   23-MAY-1997; US-047586.
PR   23-MAY-1997; US-047587.
PR   23-MAY-1997; US-047588.
PR   23-MAY-1997; US-047589.
PR   23-MAY-1997; US-047590.
PR   23-MAY-1997; US-047592.
PR   23-MAY-1997; US-047593.
PR   23-MAY-1997; US-047594.
PR   23-MAY-1997; US-047595.
PR   23-MAY-1997; US-047596.
PR   23-MAY-1997; US-047597.
PR   23-MAY-1997; US-047598.
PR   23-MAY-1997; US-047599.
PR   23-MAY-1997; US-047600.
PR   23-MAY-1997; US-047601.
PR   23-MAY-1997; US-047612.
PR   23-MAY-1997; US-047613.
PR   23-MAY-1997; US-047614.
PR   23-MAY-1997; US-047615.
PR   23-MAY-1997; US-047617.
PR   23-MAY-1997; US-047618.
PR   23-MAY-1997; US-047632.
PR   23-MAY-1997; US-047633.
PR   06-JUN-1997; US-048964.
PR   06-JUN-1997; US-048974.
PR   22-AUG-1997; US-056630.
PR   22-AUG-1997; US-056631.
PR   22-AUG-1997; US-056632.
PR   22-AUG-1997; US-056636.
PR   22-AUG-1997; US-056637.
PR   22-AUG-1997; US-056662.
PR   22-AUG-1997; US-056664.
PR   22-AUG-1997; US-056845.
PR   22-AUG-1997; US-056862.
PR   22-AUG-1997; US-056864.
PR   22-AUG-1997; US-056872.
PR   22-AUG-1997; US-056874.
PR   22-AUG-1997; US-056875.
PR   22-AUG-1997; US-056876.
PR   22-AUG-1997; US-056877.
PR   22-AUG-1997; US-056878.
PR   22-AUG-1997; US-056879.
PR   22-AUG-1997; US-056880.
PR   22-AUG-1997; US-056881.
PR   22-AUG-1997; US-056882.
PR   22-AUG-1997; US-056884.
PR   22-AUG-1997; US-056886.
PR   22-AUG-1997; US-056887.
PR   22-AUG-1997; US-056888.
PR   22-AUG-1997; US-056889.
PR   22-AUG-1997; US-056892.
PR   22-AUG-1997; US-056893.
PR   22-AUG-1997; US-056894.
PR   22-AUG-1997; US-056903.
```

```
PR   22-AUG-1997; US-056908.
PR   22-AUG-1997; US-056909.
PR   22-AUG-1997; US-056910.
PR   22-AUG-1997; US-056911.
PR   05-SEP-1997; US-057650.
PR   05-SEP-1997; US-057761.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI   Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR   WPI; 98-609887/51.
DR   P-PSDB; W75060.
PT   New isolated human genes and the secreted polypeptides they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
PS   Claim 1; Page 167; 447pp; English.
CC   This sequence represents a nucleic acid molecule which encodes a secreted
CC   human protein. The gene number, and the clone it is derived from, are
CC   detailed in the descriptor line. The gene can be used to generate fusion
CC   proteins by linking to the gene to a human immunoglobulin Fc portion
CC   (e.g. V34145) for increasing the stability of the fused protein as
CC   compared to the human protein only.
CC   The invention relates to 70 novel genes and their fragments (nucleic acid
CC   sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 70 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V34154 for described
CC   uses).
SQ   Sequence   1376 BP;   316 A;   354 C;   356 G;   347 T;

Query Match             99.1%;  Score 1277.6;  DB 1;  Length 1376;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1288;  Conservative   1;  Mismatches   0;  Indels   1;  Gaps   1;

Qy      1 AGCCAGGCGTCCCTCTGCCTGCCCACTCAGTGCAACACCCGGGAGCTGTTTGTCCTTT    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     53 AGCCAGGCGTCCCTCTGCCTGCCCACTCAGTGCAACACCCGGGAGCTGTTTGTCCTTT   112

Qy     61 GTGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 GTGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA   172

Qy    121 CCATGCAGTCGTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    173 CCATGCAGTCGTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTTGCTCATCTTTC   232

Qy    181 TGTGTGGTGCAGCCCGTTGGCAGTGTGCATCTGGGTGTCAATCGATGGGGCATCCTTTC   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 TGTGTGGTGCAGCCCGTTGGCAGTGTGCATCTGGGTGTCAATCGATGGGGCATCCTTTC   292

Qy    241 TGAAGATCTTCGGGCCACTGTGTCCAGTGCCATGGCTGGGCTACTTCC   300
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 TGAAGATCTTCGGGCCACTGTGTCCAGTGCCATGGCTGGGCTACTTCC   352

Qy    301 TCATCGGCAGCCGGCGCGTGTGGGTCTTTGCTCTTGGTTCCTGGGCTGCTATGGTGTGCTAAGA   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 TCATCGGCAGCCGGCGCGTGTGGGTCTTTGCTCTTGGTTCCTGGGCTGCTATGGTGTGCTAAGA   412

Qy    361 CTGAGAGCAAGTGTGCCCTCGTGTCGCCTGGTGTACACCACAATGGCTGAGCACTTCCTGACGTG   420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 CTGAGAGCAAGTGTGCCCTCGTGTCGCCTGGTGTACACCACAATGGCTGAGCACTTCCTGACGTG   472

Qy    421 AGGTTGCAGCTGCGTGTGGTCGCCTGGTGTACACCACAATGGCTGAGCACTTCCTGACGT   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 AGGTTGCAGCTGCGTGTGGTCGCCTGGTGTACACCACAATGGCTGAGCACTTCCTGACGT   532

Qy    481 TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT   540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT   592

Qy    541 GGAACACCACCATGAGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACCGATTTTGAGG   600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    593 GGAACACCACCATGAGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACCGATTTTGAGG   652

Qy    601 ACTCACCCTACTTCAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCA   660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 ACTCACCCTACTTCAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCA   712

Qy    661 CCAACAGCCAATGAAAACCTGCACCAAGGCTCACGACCAAAAAGTAGAGGGTT   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    713 CCAACAGCCAATGAAAACCTGCACCAAGGCTCACGACCAAAAAGTAGAGGGTT   772

Qy    721 GCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGGTGTGGCAG   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    773 GCTTCAATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGGTGTGGCAG   832

Qy    781 CTGGAATTGGGGGCCTCGAGCTGGCTGCCATGATGTGTCATGTATCTGTACTGCAATC   840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 CTGGAATTGGGGGCCTCGAGCTGGCTGCCATGATGTGTCATGTATCTGTACTGCAATC   892

Qy    841 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGC   900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    893 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGC   952

Qy    901 ACCCTGCCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA   960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    953 ACCCTGCCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA  1012

Qy    961 GAATGGACCTGCCCTTTCGTCTCCAGACTTGGGGTAGATAGGGACCACTCCTTTA-GC  1019
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   1013 GAATGGACCTGCCCTTTCGTCTCCAGACTTGGGGTAGATAGGGACCACTCCTTTANGC  1072

Qy   1020 GATGCCTGACTTCCTTCCTTCCATTGGTGGGTGATGGGTGGGGGCATTCCAGAGCCTCTAA  1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073 GATGCCTGACTTCCTTCCTTCCATTGGTGGGTGATGGGTGGGGGCATTCCAGAGCCCTCTAA  1132

Qy   1080 GGTAGCCAGTTCTGTTGCCCATTCCCCAGTCTATTAAAACCCTCATTATATGCCCCCTAGGC  1139
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 GGTAGCCAGTTCTGTTGCCCATTCCCCAGTCTATTAAAACCCTCATTATATGCCCCCTAGGC  1192

Qy   1140 CTAGTGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAAGGCATTTTATAGCCTGGGCA  1199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 CTAGTGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAAGGCATTTTATAGCCTGGGCA  1252

Qy   1200 TAAGTGAAATCAGCAGAGCCCTCGGGTGGATGTGTGAGAAGGCACTTCAAAATGCATAAAC  1259
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 TAAGTGAAATCAGCAGAGCCCTCGGGTGGATGTGTGAGAAGGCACTTCAAAATGCATAAAC  1312

Qy   1260 CTGTTACAATGTTAAAAAAAAAAAAAAAAAA  1289
          |||||||||||||||:|||||||||||||||
Db   1313 CTGTTACAATGTTAAAAAAAAAAAAAAAAAAA  1342

RESULT   4
V30916
ID   V30916 standard; cDNA; 1605 BP.
AC   V30916;
DT   14-SEP-1998 (first entry)
DE   Human secreted protein AR415_4 cDNA.
KW   AR415_4; secreted protein; protein factor; human; ds.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   CDS             437..1162
FT                   /*tag= a
FT   sig_peptide     476..514
FT                   /*tag= b
FT                   /note= "putative leader/signal peptide"
FT   mat_peptide     515..1159
FT                   /*tag= c
PN   WO9817687-A2.
PD   30-APR-1998.
```

```
PF   24-OCT-1997; U19590.
PR   24-OCT-1997; US-740274.
PR   25-OCT-1996; US-740274.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI: 98-261426/23.
DR   P-PSDB; W58380.
PT   Nucleic acid encoding secreted protein from human cells - useful,
PT   e.g. as immunomodulator, antitumour agent, promoters of tissue
PT   growth, haemostatic and thrombolytic agents etc.
PS   Claim 1; Page 66-67; 114pp; English.
CC   This cDNA clone, designated AT415_4, codes for a novel human
CC   secreted protein (see W58380). It was isolated from a human adult
CC   retina cDNA library using methods selective for cDNAs that encode
CC   secreted proteins. The clone is deposited in composite clone
CC   ATCC 98232; an oligonucleotide (see V30033) is designed to isolate
CC   the clone from the composite. The predicted AT415_4 amino acid
CC   sequence shows homology to human M35252 and CO-029 tumour associated
CC   antigens. Novel cDNA clones (see V30916-32) coding for human
CC   secreted proteins (see W58580-90) are claimed. These can be used
CC   for recombinant production of the secreted proteins for analysis,
CC   characterisation, diagnostic or therapeutic use. They can also be
CC   used as tissue or mol.wt. markers, for chromosome identification,
CC   to identify genetic disorders, to isolate new related DNA, as
CC   sources of primers for PCR, to generate antibodies, and in
CC   interaction trap assays. The secreted proteins may also have many
CC   biological activities, e.g. cytokine, immunomodulator,
CC   haematopoiesis regulating activity, tissue growth activity, activin
CC   or inhibin activity, chemotactic or chemokinetic activity,
CC   haemostatic and thrombolytic activity, receptor/ligand activity,
CC   antiinflammatory, cadherin and tumour invasion suppressor activity,
CC   and tumour inhibition activity. The proteins can be expressed in
CC   vivo from DNA, introduced in gene therapy vectors.
SQ   Sequence 1605 BP; 360 A; 427 C; 399 G; 419 T;

Query Match    98.9%;  Score 1274.8;  DB 1;  Length 1605;
Best Local Similarity  99.8%;  Pred. No. 0;
Matches 1287;  Conservative 0;  Mismatches 2;  Indels 1;  Gaps 1;

Qy    1 AGCCAGGGGTCCCTGCCTGCCACTCAGTGGCAACACCCGGAGCTGTTTTGTCCTTT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  315 AGCCAGGGGTCCCCTGCCTGCCACTCAGTGGCAACACCCGGAGCTGTTTTGTCCCTTT 374

Qy   61 GTGGAGCCCTCAGCAGTTCCCTCTTTCAGAACTCACTCCTGCCAAGAGCCCTGAACAGGAGCCA 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  375 GTGGAGCCCTCAGCAGTTCCCTCTTTCAGAACTCACTCCTGCCAAGAGCCCTGAACAGGAGCCA 434

Qy  121 CCATGCAGTGCTTCAGCTTCATTAAGACCATTAAGACCATGAATCGTCTTCAATTGCTCATCTTTC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  435 CCATGCAGTGCTTCAGCTTCATTAAGACCATTAAGACCATGAATCGTCTTCAATTGCTCATCTTTC 494

Qy  181 TGTGTGGTGCAGCCCGTTGGCAGCCCACTCGTCGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  495 TGTGTGGTGCAGCCCGTTGGCAGCCCACTCGTCGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 554

Qy  241 TGAAGATCTTCGGGCACCTGTCGTCGTCCAGTGCCATGCCAGTTTGTCAACGTTGGGCTACTTCC 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  555 TGAAGATCTTCGGGCACCTGTCGTCGTCCAGTGCCATGCCAGTTTGTCAACGTTGGGCTACTTCC 614

Qy  301 TCAATCGCAGCCGCGTTGTGGGTCTTTGCTTCTTTGGTTCCTGGGCTGCTATGGTGCTAAGA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  615 TCAATCGCAGCCGCGTTGTGGGTCTTTGCTTCTTTGGTTCCTGGGCTGCTATGGTGCTAAGA 674

Qy  361 CTGAGAGCAAGTGTGCCCTGACGTTCTTCTTCTTCATCCTCCTCCATCCTTCATTGCTG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  675 CTGAGAGCAAGTGTGCCCTGACGTTCTTCTTCTTCATCCTCCTCCATCCTTCATTGCTG 734

Qy  421 AGGTTGCAGCTGCTGGTCGCCTGGGTACACCACAATGGCTGAGCACTTCCTGACGT 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  735 AGGTTGCAGCTGCTGGTCGCCTGGGTGTACACCACAATGGCTGAGCACTTCCCTGACGT 794
```

```
Qy  481 TGCTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  795 TGCTGGTAGTGCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT 854

Qy  541 GGAACACCACATGAAAGGGCTCAAGTGCTGGGCTTCACCAACTATACGGATTTTGAGG 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  855 GGAACACCACATGAAAGGGCTCAAGTGCTGGGCTTCACCAACTATACGGATTTTGAGG 914

Qy  601 ACTCACCTACTTCAAAGAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  915 ACTCACCTACTTCAAAGAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA 974

Qy  661 CCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAGTAGAGGGTT 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  975 CCAACACAGCCAATGAAACCTGCACCAAGCAAAAGGCTCACGACCAAAAGTAGAGGGTT 1034

Qy  721 GCTTCAATCAGCTTTGTATGACATCCGAACTAATCCAGTCACCGTGGGTGTGTGCAG 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1035 GCTTCAATCAGCTTTGTATGACATCCGAACTAATCCAGTCACCGTGGGTGTGTGCAG 1094

Qy  781 CTGGAATTGGGGCCTCGAGCTGGCTGCCATGATTGTGTCCATGATCTGTACTGCCAATC 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1095 CTGGAATTGGGGCCCTCGAGCTGGCTGCCATGATTGTGTCCATGATCTGTACTGCCAATC 1154

Qy  841 TACAATAAGTCCACTCTGCCTCTGCCACTACTGCGTGCCACTACTGCTGAACTGTGAAGAGGC 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1155 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCGTGCCACTACTGCTGAACTGTGAAGAGGC 1214

Qy  901 ACCCTGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1215 ACCCTGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCA 1274

Qy  961 GAATGGACCTGCCCTTTCGTCTGCTCCAGACTTGGGCTAGATAGGGACCACTCCCTTTA-GC 1019
         |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||| ||
Db 1275 GAATGGACCTGCCCTTTCGTCTGCTCCAGACTTGGGGTAGATAGGGACCACTCCCTTTTAGGC 1334

Qy 1020 GATGCCTGACTTTCCTTCCATTGGTGGTCATTGGTGGTGGATGGGTGGGGGCATTCCAGAGCCTCTAA 1079
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1335 GATGCCTGACTTTCCTTCCATTGGTGGTCATTGGTGGTGGATGGGTGGGGGCATTCCAGAGCCTCTAA 1394

Qy 1080 GGTAGCCAGTTCGTTGCCCATTCCCCAGTCTATTAAACCCTTGATATGCCCCCTAGGC 1139
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1395 GGTAGCCAGTTCGTTGCCCATTCCCCAGTCTATTAAACCCTTGATATGCCCCCTAGGC 1454

Qy 1140 CTAGTGTGATCCCAGTGCTCTACTGGGGATGAGAGAAGGCATTTTATAGCCTGGGCA 1199
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1455 CTAGTGTGATCCCAGTGCTCTACTGGGGATGAGAGAAGGCATTTTATAGCCTGGGCA 1514

Qy 1200 TAAGTGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAAGGCACTTCAAAATGCATAAAC 1259
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1515 TAAGTGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAAGGCACTTCAAAATGCATAAAC 1574

Qy 1260 CTGTTACAAATGTTAAAAAAAAAAAAAAAAA 1289
         ||||||||||||||||||||||||||||||||
Db 1575 CTGTTACAAATGTTGAAAAAAAAAAAAAAAAAA 1604
```

```
RESULT        5
V34225
ID   V34225 standard; DNA; 1324 BP.
AC   V34225;
DT   28-JAN-1999 (first entry)
DE   Human secreted protein gene 4 clone HKCSR70.
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   development abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS   Homo sapiens.
PN   WO9833446-A2.
```

```
PD    11-SEP-1998.
PF    06-MAR-1998;  U04492.
PR    07-MAR-1997;  US-038621.
PR    07-MAR-1997;  US-040161.
PR    07-MAR-1997;  US-040162.
PR    07-MAR-1997;  US-040163.
PR    07-MAR-1997;  US-040333.
PR    07-MAR-1997;  US-040334.
PR    07-MAR-1997;  US-040336.
PR    11-APR-1997;  US-040626.
PR    11-APR-1997;  US-043311.
PR    11-APR-1997;  US-043312.
PR    11-APR-1997;  US-043313.
PR    11-APR-1997;  US-043314.
PR    11-APR-1997;  US-043315.
PR    11-APR-1997;  US-043568.
PR    11-APR-1997;  US-043569.
PR    11-APR-1997;  US-043576.
PR    11-APR-1997;  US-043578.
PR    11-APR-1997;  US-043580.
PR    11-APR-1997;  US-043669.
PR    11-APR-1997;  US-043670.
PR    11-APR-1997;  US-043671.
PR    11-APR-1997;  US-043672.
PR    11-APR-1997;  US-043674.
PR    23-MAY-1997;  US-047492.
PR    23-MAY-1997;  US-047500.
PR    23-MAY-1997;  US-047501.
PR    23-MAY-1997;  US-047502.
PR    23-MAY-1997;  US-047503.
PR    23-MAY-1997;  US-047581.
PR    23-MAY-1997;  US-047582.
PR    23-MAY-1997;  US-047583.
PR    23-MAY-1997;  US-047584.
PR    23-MAY-1997;  US-047585.
PR    23-MAY-1997;  US-047586.
PR    23-MAY-1997;  US-047587.
PR    23-MAY-1997;  US-047588.
PR    23-MAY-1997;  US-047589.
PR    23-MAY-1997;  US-047590.
PR    23-MAY-1997;  US-047592.
PR    23-MAY-1997;  US-047593.
PR    23-MAY-1997;  US-047594.
PR    23-MAY-1997;  US-047595.
PR    23-MAY-1997;  US-047596.
PR    23-MAY-1997;  US-047597.
PR    23-MAY-1997;  US-047598.
PR    23-MAY-1997;  US-047599.
PR    23-MAY-1997;  US-047600.
PR    23-MAY-1997;  US-047601.
PR    23-MAY-1997;  US-047612.
PR    23-MAY-1997;  US-047613.
PR    23-MAY-1997;  US-047614.
PR    23-MAY-1997;  US-047615.
PR    23-MAY-1997;  US-047617.
PR    23-MAY-1997;  US-047618.
PR    23-MAY-1997;  US-047632.
PR    23-MAY-1997;  US-047633.
PR    06-JUN-1997;  US-048964.
PR    06-JUN-1997;  US-048974.
PR    22-AUG-1997;  US-056630.
PR    22-AUG-1997;  US-056631.
PR    22-AUG-1997;  US-056632.
PR    22-AUG-1997;  US-056636.
PR    22-AUG-1997;  US-056637.
PR    22-AUG-1997;  US-056662.
PR    22-AUG-1997;  US-056664.
PR    22-AUG-1997;  US-056845.
PR    22-AUG-1997;  US-056862.
PR    22-AUG-1997;  US-056864.
PR    22-AUG-1997;  US-056872.
PR    22-AUG-1997;  US-056874.
PR    22-AUG-1997;  US-056875.
PR    22-AUG-1997;  US-056876.
PR    22-AUG-1997;  US-056877.
PR    22-AUG-1997;  US-056878.
PR    22-AUG-1997;  US-056879.
PR    22-AUG-1997;  US-056880.
PR    22-AUG-1997;  US-056881.
PR    22-AUG-1997;  US-056882.
PR    22-AUG-1997;  US-056884.
PR    22-AUG-1997;  US-056886.
PR    22-AUG-1997;  US-056887.
PR    22-AUG-1997;  US-056888.
PR    22-AUG-1997;  US-056889.
PR    22-AUG-1997;  US-056892.
PR    22-AUG-1997;  US-056893.
PR    22-AUG-1997;  US-056894.
PR    22-AUG-1997;  US-056903.
PR    22-AUG-1997;  US-056908.
PR    22-AUG-1997;  US-056909.
PR    22-AUG-1997;  US-056910.
PR    22-AUG-1997;  US-056911.
PR    05-SEP-1997;  US-057650.
PR    05-SEP-1997;  US-057761.
PA    (HUMA-) HUMAN GENOME SCI INC.
PI    Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI    Feng P, Ferrie AM, Fischer CL, Graves KA, Greene JM, Hu JS,
PI    Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI    Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
DR    WPI; 98-609887/51.
DR    P-PSDB; W75128.
PT    New isolated human genes and the secreted polypeptides they encode
PT    - useful for diagnosis and treatment of e.g. cancers, neurological
PT    disorders, immune diseases, inflammation or blood disorders
PS    Claim 1; Page 226; 447pp; English.
CC    This sequence represents a nucleic acid molecule which encodes a secreted
CC    human protein. The gene number, and the clone it is derived from, are
CC    detailed in the descriptor line. The gene can be used to generate fusion
CC    proteins by linking to the gene to a human immunoglobulin Fc portion
CC    (e.g. V34145) for increasing the stability of the fused protein as
CC    compared to the human protein only.
CC    The invention relates to 70 novel genes and their fragments (nucleic acid
CC    sequences: V34154-V34276; amino acid sequences W75057-W75179) which
CC    are useful for preventing, treating or ameliorating medical conditions
CC    e.g. by protein or gene therapy. Also, pathological conditions can be
CC    diagnosed by determining the amount of the new polypeptides in a sample
CC    or by determining the presence of mutations in the new polynucleotides.
CC    Specific uses are described for each of the 70 polynucleotides, based on
CC    which tissues they are most highly expressed in (see V34154 for described
CC    uses).
SQ    Sequence    1324 BP;    287 A;    344 G;    345 C;    345 T;

Query Match             98.8%;  Score 1273.6;  DB 1;  Length 1324;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1273;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        1 AGCCAGGCGTCCCTGCCTGCCTCAGTGCAACACCCGGAGCGTGTTTGTCCTTT 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       50 AGCCAGGCGTCCCTGCCTGCCTCAGTGCAACACCCGGAGCGTGTTTGTCCTTT 109

Qy       61 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      110 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 169

Qy      121 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      170 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 229

Qy      181 TGTGTGGTGCAGCCCTGTTGGCAGTGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      230 TGTGTGGTGCAGCCCTGTTGGCAGTGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 289

Qy      241 TGAAGATCTTCGGGCACTGTCGTCCATGCCAATGCAGTTGTCAACGTGGCTACTTCC 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

ID   V48113 standard; cDNA; 1288 BP.
AC   V48113;
DT   27-OCT-1998 (first entry)
DE   Nucleotide sequence encoding clone HPWAE25 of TM4SF superfamily.
KW   ds; human; receptor; immune disorder; cancers; blood disorder;
KW   juvenile rheumatoid arthritis; Graves disease.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   118..939
FT   /tag= a
FT   /product= "Clone HPWAE25"
PN   WO9831799-A2.
PD   23-JUL-1998.
PF   21-JAN-1998; U00959.
PR   21-JAN-1997; US-034205.
PR   21-JAN-1997; US-034204.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Gentz RL, Ni J, Rosen CA;
DR   WPI; 98-427559/36.
DR   P-PSDB; W61618.
PT   New isolated poly:nucleotide(s) and encoded receptor poly:peptide(s)
PT   - used to develop products for diagnosing or treating e.g. immune
PT   disorders, cancers, blood disorders or immuno-compromised disease
PT   states
PS   Claim 2; Page 36-37; 79pp; English.
CC   Clone HPWAE25 is a member of the TM4SF receptor superfamily. The
CC   products generated using the receptor can be used for treating abnormal
CC   conditions related to both an excess of and insufficient amounts of
CC   receptor activity. They can be used in the treatment of e.g. immune
CC   disorders, cancers, blood disorders, juvenile rheumatoid arthritis,
CC   Graves disease or immunocompromised disease states. The products can
CC   also be used for detection and diagnosis.
SQ   Sequence 1288 BP;   293 A;   333 C;   328 G;   334 T;

Query Match   97.9%; Score 1262; DB 1; Length 1288;
Best Local Similarity   99.8%; Pred. No. 0;
Matches 1284; Conservative   0; Mismatches   0; Indels   2; Gaps   2;

Qy   6 GGGGTCCCTGCCTGCCCACTCAGTGCAACACCCGGAGCTGTTTGTCCTTTGTGGA 65
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGGTCCCTGCCTGCCCACTCAGTGCAACACCCGGAGCTGTTTGTCCTTTGTGGA 60

Qy   66 GCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCTGAACAGGAGCCACCATG 125
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCTGAACAGGAGCCACCATG 120

Qy   126 CAGTGCTTCAGCTTCATTAAGACCATGACCAGTTGCTCAATCTTTCTGTGT 185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGTGCTTCAGCTTCATTAAGACCATGACCAGTTGCTCAATCTTTCTGTGT 180

Qy   186 GGTGCAGCCCTGTTGGCAGTTGGCGCATCTGGGTGTCAATGATGGGGCATCGTCAATGGTGCTAAGACTGAG 245
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGTGCAGCCCTGTTGGCAGTTGGCGCATCTGGGTGTCAATGATGGGGCATCGTCAATGATGGTGCTAAGACTGAG 240

Qy   246 ATCTTCGGGCCACTGTCGTCGTCCAGTGCCATGGCAGTTGTCAACGTCAGTTGTCAACGTCAGTTGTCACTTCCTCATC 305
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATCTTCGGGCCACTGTCGTCGTCCAGTGCCATGGCAGTTGTCAACGTCAGTTGTCAACGTCAGTTGTCACTTCCTCATC 300

Qy   306 GCAGCCGGCGTTGTGGTCTTGTCTCTTGCTCTGGTCAATGATGGGGCATCGTCGAG 365
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCAGCCGGCGTTGTGGTCTTGTCTCTTGCTCTGGTCAATGATGGGGCATCGTCGAG 360

Qy   366 AGCAAGTGCCCTCGTGACGTTCTTCATCCTCCTCCATCATCTTCATTGCTGAGGTT 425
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGCAAGTGCCCTCGTGACGTTCTTCATCCTCCTCCATCATCTTCATTGCTGAGGTT 420

Qy   426 GCAGCTGCTGTGGTCGCCTTGGTGTCACGCTGTACACCACAATGGCTGAGCACTTCCTGACGTTGCTG 485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GCAGCTGCTGTGGTCGCCTTGGTGTCACGCTGTACACCACAATGGCTGAGCACTTCCTGACGTTGCTG 480

Qy   486 GTAGTGCCTGCCATCAAGAAAGATTATGGTTCACTCAAGTGTGGAAC 545
           ||||||||||||||||||||||||||||||||||||||||||||||||

Db   290 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCCATGCCAGTGCAGTTGTCAAGCTGGCTACTTCC 349
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301 TCATCGCAGCCGGCGTTGTGGTCTTGTCTTGCTCTTGCTCTTGGTCTTGGGCTGCTATGGTGCTAAGA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   350 TCATCGCAGCCGGCGTTGTGGTCTTGTCTTGCTCTTGCTCTTGGTCTTGGGCTGCTATGGTGCTAAGA 409

Qy   361 CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   410 CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCATCTTCATTGCTG 469

Qy   421 AGGTTGCAGCTGCTGTGGTCGCCTTGGTGTACACCACAATGGCTGAGCACTTCCTGACGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   470 AGGTTGCAGCTGCTGTGGTCGCCTTGGTGTACACCACAATGGCTGAGCACTTCCTGACGT 529

Qy   481 TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   530 TGCTGGTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGT 589

Qy   541 GGAACACCACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGG 600
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   590 GGAACACACCATGAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGG 649

Qy   601 ACTCACCCTACTTCAAAGAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   650 ACTCACCCTACTTCAAAGAGAACAGTGCCTTTCCCCATTCTGTTGCAATGACAACGTCA 709

Qy   661 CCAACAGCCAATGAAACCTGCACCACAGCTCACGACCAAAAAGTAGAGGGTT 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   710 CCAACAACAGCCAATGAAACCTGCACCACAGCTCACGACCAAAAAGTAGAGGGTT 769

Qy   721 GCTTCAATCAGCTTTGTATGACAATCAGCTGCCAACTAATGCAGTCACCGGTGGTGTGGCAG 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   770 GCTTCAATCAGCTTTGTATGACAATCAGCTGCCAACTAATGCAGTCACCGGTGGTGTGGCAG 829

Qy   781 CTGGAATTGGGGGCCTCGAGCTGCCTGCGCTGCCATGAGTGTCCATGTATCGTACTGCAATC 840
           |||||||||||||||||||||||||||||||||||::|||||||||||||||||||||||||
Db   830 CTGGAATTGGGGGCCTCGAGCTGCCTGCGCTGCCATGAGTGTKTCCATGTATCGTACTGCAATC 889

Qy   841 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCCACTAGTGGGAACTGTGAAGAGGC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   890 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCCACTAGTGGGAACTGTGAAGAGGC 949

Qy   901 ACCCTGGCAAGCAGCAGCAGTGATTGGGGGACAGGAGTCAACAATGTCACTTGGGCCA 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   950 ACCCTGGCAAGCAGCAGCAGTGATTGGGGGACAGGAGTCAACAATGTCACTTGGGCCA 1009

Qy   961 GAATGGACCTGCCCTTTCTGCTCCAGACTGGGGCTAGATAGGGACCACTCCTTTAGCG 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 GAATGGACCTGCCCTTTCTGCTCCAGACTGGGGCTAGATAGGGACCACTCCTTTAGCG 1069

Qy   1021 ATGCCTGACTTTCCTTCCATTGGTGGATGGGTGGGGGCCATTCCAGAGCCTCTAAG 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1070 ATGCCTGACTTTGCTTCTTCCATTGGTGGATGGGTGGGGGCCATTCCAGAGCCTCTAAG 1129

Qy   1081 GTAGCCAGGTCTGTTGCCCAGTCGTCATTAAACCCTTGATATATGCCCCCTAGGCC 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1130 GTAGCCAGTTCTGTTGCCCATTCCCCAGTCGTTATTAAACCCTTGATATGCCCCCTAGGCC 1189

Qy   1141 TAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1190 TAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCAT 1249

Qy   1201 AAGTGAAATCAGCAGAGCCTCTGGGTGATGTGTAGAGGTGTAGAGGCACTTCAAAATGCATAAACC 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1250 AAGTGAAATCAGCAGAGCCTCTGGGTGATGTGTAGAGTGTGTAGAGGCACTTCAAAATGCATAAACC 1309

Qy   1261 TGTTACAATGTTAAA 1275
           |||||||||||||||
Db   1310 TGTTACAATGTTAAA 1324

RESULT   6
V48113

PA   (INCY-), INCYTE PHARM INC.
PI   Goli SK, Hillman JL;
DR   WPI; 98-481208/41.
PT   Human tumour-associated antigen PRAT - useful for stimulating cell
PT   proliferation and screening for antagonists useful to treat or
PS   prevent cell proliferation disorders e.g. cancers
PS   Disclosure; Fig 1A-1C; 54p; English.
CC   This is the nucleotide sequence of the human tumour-associated antigen
CC   (PRAT) used in the method of the invention for stimulating cell
CC   proliferation and screening for antagonists useful to treat or prevent
CC   cell proliferation disorders such as cancer, and genetic defect e.g.
CC   sickle cell anaemia. The polypeptides can be combined with a suitable
CC   carrier in pharmaceutical compositions, and also used to screen for
CC   antagonists, agonists, and to generate antibodies. PRAT agonists can
CC   be added to a cell to stimulate cell proliferation by increasing or
CC   prolonging the activity of PRAT as above. The antagonists can be
CC   combined with a suitable carrier in pharmaceutical compositions, which
CC   can be administered to subjects to treat or prevent disorders
CC   associated with cell proliferation, especially cancers. Antibodies
CC   specific for PRAT may be used directly as antagonists, or indirectly
CC   as a targeting or delivery mechanism to bring pharmaceutical agents to
CC   PRAT-expressing cells. They are also useful to diagnose conditions or
CC   diseases characterised by PRAT expression and to monitor therapeutic
CC   interventions. The polynucleotide encoding PRAT, or complementary
CC   sequences, can be used to produce hybridisation probes, useful to
CC   detect polynucleotides or fragments encoding PRAT, e.g. to diagnose
CC   diseases relating to polypeptide expression or monitor PRAT regulation
CC   during therapeutic intervention.
SQ   Sequence 933 BP;   198 A;   254 C;   236 G;   245 T;

Query Match   72.0%;   Score 928;   DB 1; Length 933;
Best Local Similarity   100.0%; Pred. No. 2.1e-266;
Matches 928;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 AGCCAGGCGTCCTCTGCCTCCAGTCGCCACTCAGTGCCAACACCCGGGAGCTGTTTGTCCTTT   60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 AGCCAGGCGTCCTCTGCCTCCAGTCGCCACTCAGTGCCAACACCCGGGAGCTGTTTGTCCTTT   65

Qy   61 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA   120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA   125

Qy   121 CCATGCAGTGCTTCAGCTTCATTAAGACCATGAGATCCTCTTCAATTGCTCATCTTTC   180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 CCATGCAGTGCTTCAGCTTCATTAAGACCATGAGATCCTCTTCAATTGCTCATCTTTC   185

Qy   181 TGTGTGTGCAGCCCTGTTGGCAGTGGCATCGGGTGTCAATCGATGGGGCATCCTTTC   240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186 TGTGTGTGCAGCCCTGTTGGCAGTGGCATCGGGTGTCAATCGATGGGGCATCCTTTC   245

Qy   241 TGAAGATCTTCGGCCACTGCGTCGTCCAGTGCCATGCAGTTGTCAACGTGCTACTTCC   300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246 TGAAGATCTTCGGCCACTGCGTCGTCCAGTGCCATGCAGTTGTCAACGTGCTACTTCC   305

Qy   301 TCATCGCAGCCGCCGCTGCCTGCCGTCCTTGTCTTGCTCTTCGGACGTGTTTGTCCTTT   360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306 TCATCCAGCCGCCGGTTGCCTGCCGTGGTCCTTGTCTTGCTCTTCGGACGTGTTTGTCCTTT   365

Qy   361 CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCAGAACTCACTTCATTGCTG   420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   366 CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCAGAACTCACTTCATTGCTG   425

Qy   421 AGGTTGCAGCTGCTGTGGTCCGCCTTGGTGTACCACCAATGCTGAGCACTTCCTGACGT   480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   426 AGGTTGCAGCTGCTGTGGTCCGCCTTGGTGTACCACCAATGCTGAGCACTTCCTGACGT   485

Qy   481 TGCTGGTAGTGCCTGCCGCCATCAAGAAAGATATGGTTCCCAGGAAGACTTCACTCAAGTGT   540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   486 TGCTGGTAGTGCCTGCCGCCATCAAGAAAGATATGGTTCCCAGGAAGACTTCACTCAAGTGT   545

Qy   541 GGAACACCACCATGAAAGGGCTCAAGTGCGTGGCTCACCAACTATACGTGATTTGAGG   600

Db   481 GTAGTGCCTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAAC   540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   546 ACCACCAGAAAGGGCTCAAGTGCTCGTGTGTCCAACTATACGGATTTGAGGACTCA   605

Db   541 CCCTACTTCAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCAAC   600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   601 CCCTACTTCAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCAAC   660

Db   666 ACAGGCAATGAAAACCTGCACCAAGCAAAAGCTCACGACCAAAAAAGTAGAGGGTTGCTTC   725
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   661 ACAGGCAATGAAAACCTGCACCAAGCAAAAGCTCACGACCAAAAAAGTAGAGGGTTGCTTC   720

Db   726 AATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCAGTGGTGGTGTGGCAGCTGGA   785
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   721 AATCAGCTTTGTATGACATCCGAACTAATGCAGTCAGTCAGTGGTGGTGTGGCAGCTGGA   780

Db   786 ATTGGGGGCCTCGAGCTGCTGCCATG-ATTGTGTCCATGTATCTGTACTGCAATCTACA   844
             |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Qy   781 ATTGGGGGCCTCGAGCTGCTGCCATGTATTGTGTCCATGTATCTGTACTGCAATCTACA   840

Db   845 ATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGAACTGTGAAGAGGCACCC   904
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   841 ATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGAACTGTGAAGAGGCACCC   900

Db   905 TGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCAGAAT   964
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   901 TGGCAAGCAGCAGTGATTGGGGAGGGGACAGGATCTAACAATGTCACTTGGGCCAGAAT   960

Db   965 GGACCTGCCCTTTCTGCTCCAGACTTGGGCTCCAGAGGGCTAGATAGGGACCACTCCTTTA-GCGATG   1023
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   961 GGACCTGCCCTTTCTGCTCCAGACTTGGGCTGGTGATGGGTGGGGGCATTCCAGAGCCTCTAAGGTG   1020

Db   1024 CCTGACTTTCCTTCCTTCCATTGGTGGGTTGGATGGGTGGGGGGCATTCCAGAGCCTCTAAGGTA   1083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 CCTGACTTTCCTTCCTTCCATTGGTGGGTTGGATGGGTGGGGGGCATTCCAGAGCCTCTAAGGTA   1080

Db   1084 GCCAGTTCGTTGCCCATTCCCCAGTCTCATTAAACCCTTGATATGCCCCTAGGCCTAG   1143
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 GCCAGTTCGTTGCCCATTCCCCAGTCTCATTAAACCCTTGATATGCCCCTAGGCCTAG   1140

Db   1144 TGGTGATCCCAGTGCTCTACTGTGCTCTACTGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCATAAG   1203
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 TGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTTATAGCCTGGGCATAAG   1200

Db   1204 TGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAGGCACTCAAAATGCATAAAACCTGT   1263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1201 TGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAGGCACTCAAAATGCATAAAACCTGT   1260

Db   1264 TACAATGTTAAAAAAAAAAAAAAAA 1289
             |||||||||||||||||||||
Qy   1261 TACAATGTTAAAAAAAAAAAAAAAA 1286

RESULT   7
V54014
ID   V54014 standard; cDNA; 933 BP.
AC   V54014;
DT   02-DEC-1998 (first entry)
DE   Nucleotide sequence of the human tumour-associated antigen.
KW   Human; tumour-associated antigen; PRAT; stimulation; cell proliferation;
KW   antagonist; cancer; genetic defect; sickle cell anaemia; agonist;
KW   antibody; hybridisation; probe; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   128..855
FT   /*tag= a
FT   /product= "human tumour-associated antigen"
PN   WO9838310-A1.
PD   03-SEP-1998.
PF   27-FEB-1998; U03953.
PR   28-FEB-1997; US-808148.

```
Db       |||||||||||||||||||||||||||||||||||||||||||
     546 GGAACACCACCATGAAAGGCTCAAGTGCTGTGGCTTCCCCATTCTGTTGCAATGACAACGTCA 605

Qy   601 ACTCACCCTACTTCAAAGAGAACAGCAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCA 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   606 ACTCACCCTACTTCAAAGAGAACAGCAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCA 665

Qy   661 CCAACACAGCCAATGAAACCTGCACCAAGCAAAGGCTCACGACCAAAAAGTAGAGGGT 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   666 CCAACACAGCCAATGAAACCTGCACCAAGCAAAGGCTCACGACCAAAAAGTAGAGGGT 725

Qy   721 GCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGGTGGCAG 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   726 GCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGGTGGCAG 785

Qy   781 CTGGAATTGGGGGGCCTCGAGCTGGCTGCCGCCATGATTGTGTGTCCACATGTGTACTGCAATC 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   786 CTGGAATTGGGGGGCCTCGAGCTGGCTGCCGCCATGATTGTGTGTCCACATGTGTACTGCAATC 845

Qy   841 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGC 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   846 TACAATAAGTCCACTTCTGCCTCTGCCACTACTGCTGCCACATGGGAACTGTGAAGAGGC 905

Qy   901 ACCCTGGCAAGCAGCAGTGATTGGGGGA 928
         |||||||||||||||||||||||||||
Db   906 ACCCTGGCAAGCAGCAGTGATTGGGGGA 933
```

RESULT 8

```
V58493/c
ID   V58493;  standard; cDNA; 801 BP.
AC   V58493;
DT   08-DEC-1998  (first entry)
DE   3' fragment of prostate tumour specific gene J1-19.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-805596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
CC   Claim 1; Page 42; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 801 BP; 209 A; 208 C; 198 G; 164 T;
```

```
Query Match     46.7%;  Score 602.4; DB 1; Length 801;
Best Local Similarity  91.7%;  Pred. No. 1.5e-169;
Matches  714;  Conservative   0;  Mismatches  55;  Indels  10;  Gaps  8;

Qy    23 CCACTCAGTGCAACACCCGGCAGCTGTTTGTCCTTTGTGGAGCCTCAGCAGTTCCCTC 82
         |||  |||||||||||||||||||||||||||||  ||  || |||||
Db   779 CCAATTCAGGGCAACACCCGCGGGGTGTTTTNTCNTTTGGGGGGCTNANGCNGTTCCTTT 720

Qy    83 TT-TCAGAACTCACTGCCAAGAGCCTGAACAGGAGCCACCATGCAGTGCTTCAGCTTCA 141
         || |||||||||||||||||||||||||||| |||||||||  || ||||
Db   719 TTNTCGAATTCACTGCCAAGAGCCCTGAACAGGGGCCACCATGCANGGTTTCAG-TTCA 661
```

Right column:

```
Qy   142 TTAAGACCATGATGATCCTCTTCAATTGCTCCA-TCTTTCGTGTGTGTGCAGCCCTGTTG 200
         ||||||| || |||| | |||||||||||||  | ||  | |||||||||||||||||||
Db   660 TTAAGACCANGANGATCCTTCTTCAATGGCTCATTCTTTCNGNGGNGGCAGCCCTGTNG 601

Qy   201 GCAGTGGCATCGGGTG-TCAATCGATGGGCATCC--TTTCTGAAGATCTTCGGGC- 257
         |||||||||||||| | ||||||||||||||||  | ||  |||||||||||||
Db   600 GCAGTGGCATNTGGGTGTTCAATCGATGGGCATCCTTTTCTGAAGATNTTCGGGCCA 541

Qy   257 ACTGTCGTCCAGT-GCCATGCAGTTGTCAACGTGGGCTACTTCCT--CATCGCAGCCGG 313
         |||||||| || |||||||||||||||||||| | ||||| |||  ||||||||| |
Db   540 ACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTANTTCTTTCATTCGCAGCNGG 481

Qy   314 CGTTGTGGTCTTGCCTGGTTCCTGGGCTGCTGCCATGGTGCTAAGACTGAGAGCAAGTG 373
         |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   480 CGTTGTGGTCTTGCCTGGTTCCTGGGCTGCTGCCATGGTGCTAAGACGAGAGCAAGTG 421

Qy   374 TGCCCTCGACGTTCTTCTCATCCTCCTCATCTTCATTGCTGAGGTTGCAGCTGC 433
         ||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   420 TGCCCTCGNGACGTTCTTCTCATCCTCCTCATCTTCATTGCTGAGGTTGCAGCTGC 361

Qy   434 TGTGGTCGCCTTGGTGTACACCACCAATGGCTGAGCACTTCCTGACCTTGCTTGGTGTAGTGCC 493
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   360 TGTGGTCGCCTTGGTGTACACCACCAATGGCTGAGCACTTCCTGACGTTGCTTGGTAGTGCC 301

Qy   494 TGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACCAT 553
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   300 TGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACCAT 241

Qy   554 GAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTT 613
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   240 GAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTT 181

Qy   614 CAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCACACAGCCAA 673
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180 CAAAGAGAACAGTGCCTTCCCCATTCTGTTGCAATGACAACGTCACCACACAGCCAA 121

Qy   674 TGAAACCTGCACCAAGCAAAGGCTCACGACCAAAAAGTAGAGGGTTGCTTCAATCAGCT 733
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   120 TGAAACCTGCACCAAGCAAAGGCTCACGACCAAAAAGTAGAGGGTTGCTTCAATCAGCT 61

Qy   734 TTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGGTGGCAGCTGGAATTGGG 792
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    60 TTTGTATGACATCCGAACTAATGCAGTCAGTCACCGTGGGTGGTGGCAGCTGGAATTGGG 2
```

RESULT 9

```
V61150/c
ID   V61150;  standard; cDNA; 801 BP.
AC   V61150;
DT   06-JAN-1999  (first entry)
DE   3' cDNA sequence of prostate tumour clone J1-19.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609986/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 41; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 801 BP; 209 A; 208 C; 198 G; 164 T;
```

PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1: Page 42; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence   740 BP;   161 A;   200 C;   160 G;   194 T;

Query Match   44.3%; Score 570.8; DB 1; Length 740;
Best Local Similarity   89.7%; Pred. No. 3.6e-160;
Matches 647; Conservative   0; Mismatches   63; Indels   11; Gaps   4;

```
Qy    1 AGCCAGGGTCCCTCGCTGCCACTCAGTGCAACACCCGGAGCTGTTTTGTCCTTT    60
          ||||||| |||||||||||||||||||||| |||||||||| ||||||||||||||
Db    6 AGCCAGGCGTCCCTCGCTGCCACTCAGTGCCAACACCCGGGAGCTGTTTTGTCCTTT    65

Qy   61 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA   120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA   125

Qy  121 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  126 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC   185

Qy  181 TGTGTGGTGCAGCCCTGTTGGCAGTGGCAGTGTCAATCGATGGGGCATCCTTTC       240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  186 TGTGTGGTGCAGCCCTGTTGGCAGTGGCAGTGTCAATCGATGGGGCATCCTTTC       245

Qy  241 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  246 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC   305

Qy  301 TCATCGCAGCCGGCGGTTGTGGTCTTTGCTCTTGCTCCTGGGCTGTATGGTGCTAAGA   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  306 TCATCGCAGCCGGCGGTTGTGGTCTTTGCTCTTGCTCCTGGGCTGTATGGTGCTAAGA   365

Qy  361 CTGAGAGCAAGTGTGCCCCGTGACGTTCTTCTCATCCTCCTCCTCCTCATTGCTG     420
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  366 CGGAGAGCAAGTGTGCCCCGTGACGTTCTTCTCATCCTCCTCCTCCTCATTGCTG     425

Qy  421 AGGTTGCAGCTGCTGTGCTGCGTGCCTTGGTGTGCACCACCATTCCTGACGT       480
          |||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  426 AGTTGCAGCTGCTGTGCTGCGTGCCTTGGTGTGCACCACCATTCCTGACGT       485

Qy  481 TGCTGGTAGTGCCTGCCACCATGAAAGATTAT-GGTTCCCAGGAAGACTTCACTCAAGTG   539
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  486 TGCTGGTANTGCCTGCCATCAANAAGATTATGGTTCCCAGGAAAAATTCACTCAANTN   545

Qy  540 TGGAACAACCATGAAAGGG---CTCAAGTGCTGTGCTTCACCAACTATACGG---AT     593
          ||||||||||| ||||||     |   || ||||||||  |||||||||||||      ||
Db  546 TGGAACACCNCATGAAAAGGGCTCCAATTTCTGNTGCTTCCCAACTATACCGGAATT   605

Qy  594 TTTGAGGACTCACCCTACTTCAAAGAGAACAGTG---CCTTTCCCCATTCGTGTTGCAA   649
          || |||| || |||||||||||||| |||||||     ||||||||||    |||||||
Db  606 TTGAAAGANTCNCCCTACTTCCAAAAAAAAANNTTGCTTTNCCCCNTTNCCCNTTGTTGCAA   665

Qy  650 TGACAAGTCACCAACACAGCCAATGAAACCTGCACCAAGGCAAAGGTCACGACCAAAA   709
          |||||||| |||||||| ||  ||||||| ||||||  ||||  ||||||||||||||
Db  666 TGAAACNTCCCAANACNGCCAATNAAAACCTGCCCNNNCAAAAAAGGNTCNCAACAAAA   725

Qy  710 A 710
          |
Db  726 A 726
```

Query Match   46.7%; Score 602.4; DB 1; Length 801;
Best Local Similarity   91.7%; Pred. No. 1.5e-169;
Matches 714; Conservative   0; Mismatches   55; Indels   10; Gaps   8;

```
Qy   23 CCACTCAGTGCAACACCCGGGAGCTGTTTTGTCCTTTGTTGGAGCCTCAGCAGTTCCCTC    82
          ||||||||||||||||| ||||||   |||      ||  ||||||    ||||  |||
Db  779 CCAATTCAGGGCAACACCCGGGGGTGTTTTNTCNTTTGGGGGGGCCTNANGCNGTTCCTTT   720

Qy   83 TT-TCAGAACTCACTGCCAAGAGCCAGCCACCATGCAGTGCTTCAGCTTCA          141
          || |||||||||||||||||||||||||||||||||||||||||||||||||
Db  719 TTNTCAGAATTCACTGCCAAGAGCCAGCCACCATGCANGTTCAG-TTCA          661

Qy  142 TTAAGACCATGATGATCCTCTTCAATTGCTCA-TCTTTCTGTGTGGTGCAGCCCTGTTG   200
          ||||||||||| |||| ||||||||||||||| |||||||||||||||||||||||||||
Db  660 TTAAGACCANGANGATCCTTCAATGGCCTCATTCTTTCNCNGNGNGGNGCAGCCCTGTNG   601

Qy  201 GCAGTGGCAGTCATCTGGGTG-TCAATCGATGGGGCATCC--TTTCTGAAGATCTTCGGGCC-   257
          ||||||||||||||||||||| ||||| ||||||||||||    ||||||||||||||||||
Db  600 GCAGTGGGCATNGGGTGTTCAATCGATNGGGGCATCCCTTTTCGGCAGTTNTCGGGCCA    541

Qy  257 ACTGTCGTCCAGT-GCCATGCAGTTGTCAACGTGGGCTACTTCCT--CATCGCAGCCGG   313
          ||||||||||||| ||||||||||||||||||||||||||||||    |||||||||||
Db  540 ACTGTCGTCCAGTGGCCATGCAGTTGTCAACGTGGGCTANTTCTTCNCAGCGCAGCNGG   481

Qy  314 CGTTGTGGTCTTTGCTCTTGGCTGCTGTATGGTGCTAAGACTGAGAGCAAGTG       373
          ||||| ||||||||||||| |||||||||||||||||||||| |||||||||||
Db  480 CGTTGNGGTCTTTGCTCTTGNTGCTGTATGGTGCTAAGACGGAGAGCAAGTG       421

Qy  374 TGCCCTCGTGACGTTCTTCTCATCCTCCTCCTCATTGCTGCTGAGGTTGCAGCTGC   433
          ||||| ||| || ||||||||||||||||||| |||||||||||||||||||||||
Db  420 TGCCCTCGNGACGTTCTTCTCATCCTCCCCATCTTCATTGCTGCTGAGGTTGCAGCTGC   361

Qy  434 TGTGGTCGCCTTGGTGTGTACACCACCAATGGCTGCACTTCCTGACGTTGCTGGTAGTGCC   493
          ||||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  360 TGTGGTCGCCTTGGTGTGTACACCACCAATGGCTGCACTTCCTGACGTTGCTGGTAGTGCC   301

Qy  494 TGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACCAT   553
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  300 TGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACCAT   241

Qy  554 GAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTT   613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  240 GAAAGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATTTTGAGGACTCACCCTACTT   181

Qy  614 CAAAGAGAACAGTGCCTTTCCCCATTCGTGTTGCAATGACAACGTCACCAACACAGCCAA   673
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  180 CAAAGAGAACAGTGCCTTTCCCCATTCGTGTTGCAATGACAACGTCACCAACACAGCCAA   121

Qy  674 TGAAACCTGCACCAAGGCAAAGCCACCGTCACGACCAAAAAGTAGAGGGTTGCTTCAATCAGCT   733
          ||||||||||| |||| |||||||||||||||||||||||||||||||||||||||||||
Db  120 TGAAACCTGCAACCAAGGCAAAGCCACCGTCACGACCAAAAAGTAGAGGGTTGCTTCAATCAGCT    61

Qy  734 TTTGTATGACATCCGAACTAATGCAGTCAGTCAGTCACCGTGGTGGTGTGGCAGCTGAATTGGGG   792
          ||||| |||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db   60 TTTGTATGACATCCGAACTAATGCAGTCAGTCAGTCACCGTGGGTGGTGTGGCAGCTGAATTGGGG     2
```

RESULT 10
V58494
ID   V58494 standard; cDNA; 740 BP.
AC   V58494;
DT   08-DEC-1998 (first entry)
DE   5' fragment of prostate tumour specific gene J1-19.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PF   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-805596.
PR   01-AUG-1997; US-904809.

```
RESULT   11
V61151
ID   V61151 standard; cDNA; 740 BP.
AC   V61151;
DT   06-JAN-1999 (first entry)
DE   5' cDNA sequence of prostate tumour clone J1-19.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 41-42; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence    740 BP;    161 A;    200 C;    160 G;    194 T;

Query Match          44.3%;  Score 570.8;  DB 1;  Length 740;
Best Local Similarity 89.7%;  Pred. No. 3.6e-160;
Matches  647;  Conservative   0;  Mismatches  63;  Indels  11;  Gaps   4;

Qy     1 AGCCAGGCGTCCCTCGCCTGCCACTCAGTGCAACACCCGGGAGCTGTTTTGTCCTTT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6 AGCCAGGCGTCCCTCGCCTGCCACTCAGTGCAACACCCGGGAGCTGTTTTGTCCTTT 65

Qy    61 GTGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGGCCA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66 GTGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGGCCA 125

Qy   121 CCATGCAGTGTTCAGCTTCATTAAGACCATGAGATCCTCTTCAATTGCTCATCTTTC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126 CCATGCAGTGTTCAGCTTCATTAAGACCATGAGATCCTCTTCAATTGCTCATCTTTC 185

Qy   181 TGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCC 240
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   186 TGTGTGGTGCAGCCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCC 245

Qy   241 TGAAGATCTTCGGGCCACTGTCGTCCAGTTGTCAACGTGGGCTACTTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246 TGAAGATCTTCGGGCCACTGTCGTCCAGTTGTCAACGTGGGCTACTTCC 305

Qy   301 TCATCCGAGCCGGCGGTGTGTGTCTCTATGGTTCTGGGCTGCTATGGTGCTAAGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306 TCATCCGAGCCGGCGGTGTGTGTCTCTATGGTTCTGGGCTGCTATGGTGCTAAGA 365

Qy   361 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCCATCCTCCTCCTCATCTTCATTGCTG 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   366 CGGAGAGCAAGTGTGCCCTCGTGACGTTCTTCTTCCATCCTCCTCCTCATCTTCATTGCTG 425

Qy   421 AGGTTGCAGTGCTTGCTGTGGTGTACACCACCAAATGGCTGAGCACTTCCTGACGT 480
            ||||||||||||||||||||||||||||||||||||||||| | ||||||||||||
Db   426 AAGTTGCAGTGCTGCTGTGGTGTACACCACCAAATGGCTGAACCATTCCTGACGT 485

Qy   481 TGCTGGTAGTGCCTGCCATCAAGAAAGATTAT-GGTTCCCAGGAGACTTCACTCAAGTG 539
           ||||||| ||||||||||||||||||||||| |||||||||| | |||||||||||
Db   486 TGCTGGTANTGCCTGCCATCAANAAAGATTATGGTTCCCAGGAAAAATTCACTCAANTN 545

Qy   540 TGGAACACCACCATGAAAGGG---CTCAAGTGCTGTGGCTTCACCAACTATACGG---AT 593
           ||||||| ||||||| ||||||| ||  || ||||| |   |   ||||
```

```
Db   546 TGGAACACCNCCATGAAAAGGGCTCCAATTCTGTGNTGGCTTCCCAACTATACCGGAATT 605

Qy   594 TTTGAGGACTCACCCTACTTCAAAGAGAACAGTG----CCTTCCCCATTCGTGTTGCAA 649
           |||| || || ||  ||  ||||| || |  |||    || | ||| | | | |||||
Db   606 TTGAAAGANTCNCCCTACTTCCAAAAAAAANANNTTGCCCTTTNCCCCNTTCGTGTTGCAA 665

Qy   650 TGCACACGTCACCAACACAGCCAATGAAACCTGCACCAAGGCTCACGACCAAAA 709
           |||| ||  ||  ||  || ||   ||||     ||  |||| | | | ||||||
Db   666 TGAAAACNTCCCAANACNGCCAATNAAACCTGCCNNNCAAAGAAGNTCNCAAACAAAA 725

Qy   710 A 710
           |
Db   726 A 726
```

```
RESULT   12
V58490
ID   V58490 standard; cDNA; 729 BP.
AC   V58490;
DT   08-DEC-1998 (first entry)
DE   5' fragment of prostate tumour specific gene N1-1862.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03690.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-806596.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 40-41; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence    729 BP;    157 A;    203 C;    165 G;    191 T;

Query Match          43.5%;  Score 560.8;  DB 1;  Length 729;
Best Local Similarity 94.9%;  Pred. No. 3.3e-157;
Matches  650;  Conservative   0;  Mismatches  26;  Indels   9;  Gaps   7;

Qy     1 AGCCAGGCGTCCCTCGCCTGCCACTCAGTGCAACACCCGGGAGCTGTTTTGTCCTTT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2 AGCCAGGCGTCCCTCGCCTGCCACTCAGTGCAACACCCGGGAGCTGTTTTGTCCTTT 61

Qy    61 GTGGAGCCTCAGCAGTTCCCTCTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGGCCA 120
           |||| ||||||||||||||||||||||||||| ||||||||| ||||||||||||||||||
Db    62 GTGGANCCTCAGCAGTNCCCTCTTTCAGAACTCAGTGCCAAGANCCCTGAACAGGAGGCCA 121

Qy   121 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 181

Qy   181 TGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 TGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 241

Qy   241 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCGTCAACGTGGGCTACTTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCGTCAACGTGGGCTACTTCC 301

Qy   301 TCATCGGACCGGCCGGTTGTGGTTCCTTGCTCTTGGTTCCTGGGCTGCTATGTGCTTGGTCTATGTGCTGTATGTGCTCTATGTGCTAAGA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TCATCGGACCGGCCGGTTGTGGTTCCTTGCTCTTGGTTCCTGGGCTGCTATGTGCTAAGA 360
```

```
Qy    181  TGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTC  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182  TGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTC 241

Qy    241  TGAAGATCTTCGGGCCACTGCGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242  TGAAGATCTTCGGGCCACTGCGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC 301

Qy    301  TCATCGCAGCCGGCGTTGTGGTCTTGCTTGCTCTTCCTGGGCTGCTATGGTGCTAAGA    360
           ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    302  TCATCGCAGCCGGCGTTGTGGTCTTGCTTGCTCTAGGTTCCTGGGCTGCTATGGTGCTAAGA 361

Qy    361  CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362  CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG   421

Qy    421  AGGTTGCAGCTGCTGTGTGTCGCCTTGGTGTACACCACCAAATGGCTGAGCACTTCCTGACGT 480
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422  AGGTTGCA-ATGCTGTGTGTCGCCTTGGTGTACACCACCAAATGGCTGAGCACTTCCTGACGT 480

Qy    481  TGCTGGTAGTGCCTGCCATCAAGAAA--GATTATGGTTCCCAGGAAGACTTCACTCAAGT 538
           |||||||| |||||||||||||||||| ||  ||||||||||||||||||||||||||||
Db    481  TGCTGGTAATGCCTGCCATCAANAAAAAGATTATGGTTCCCAGGAAANACTTCACTCAAGT 540

Qy    539  G-TGGAACACCACCATGAAAAGGGCTCAAGTGCTGTGGCTTC-ACCAACTATACGGGATTT 596
           | ||||||||||||||||| |||||||||||||||||||| | |||||||||||||||||
Db    541  GTTGGAACACCACCATGAAAAGGGCTCAAGTGCTGTGGCTTCNNCCAACTATACGGGATTTT 600

Qy    597  GAGGACTTCACCCTACTTCAAAGAACAGTGCCTTCCCCCA-TTCTGTTGCAAT--GAC   653
           || ||| |||||||||||||||||| || |||||||||||| |||||||||||  ||||
Db    601  GAAGANTCA-CCTACTTCAAAGAAAAGTGCCTTTCCCCCATTTCTGTTGCAATTGACA  659

Qy    654  AACGTCACCAACACAGCCAATGAAA 678
           |||||||||||||||||||||||||
Db    660  AACGTCCCCAACACAGCCAATTGAA 684
```

RESULT 14
V58489/C
ID   V58489 standard; cDNA; 751 BP.
AC   V58489;
DT   08-DEC-1998 (first entry)
DE   3' fragment of prostate tumour specific gene N1-1862.
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy; ss.
OS   Homo sapiens.
PN   WO9837418-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03590.
PR   09-FEB-1998; US-904809.
PR   25-FEB-1997; US-805696.
PR   01-AUG-1997; US-904809.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-480805/41.
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
PS   Claim 1; Page 40; 141pp; English.
CC   This sequence represents a human prostate tumour specific gene, and can
CC   be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   encoded by this sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
SQ   Sequence 751 BP; 188 A; 196 C; 167 G; 156 T;

```
Query Match         41.0%;  Score 528.4;  DB 1;  Length 751;
Best Local Similarity  88.0%;  Pred. No. 1.5e-147;
Matches  654;  Conservative  0;  Mismatches  75;  Indels  14;  Gaps  9;
```

```
Db    302  TCATCGCAGCCGGCGTTGTGGTCTTGCTTGCTCTTAGGTTCCTGGGCTGCTATGGTGCTAAGA 361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361  CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362  CTGAGAGCAAGTGCCCTCGTGACGTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTG   421
Qy    421  AGGTTGCAGCTGCTGTGGTCGCCCTTGGTGTACACCACCACCACCAAATGGCTGAGCACTTCCTGACGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422  AGGTTGCA-ATGCTGTGGTCGCCCTTGGTGTACACCACCACCAAATGGCTGAGCACTTCCTGACGT 480
Qy    481  TGCTGGTAGTGCCTGCCATCAAGAAA--GATTATGGTTCCCAGGAAGACTCCACTCAAGT 538
           |||||||| |||||||||||||||||| ||  |||||||||||||||||||||||||||
Db    481  TGCTGGTAATGCCTGCCATCAANAAAAAGATTATGGTTCCCAGGAAANACTTCACTCAAGT 540
Qy    539  G-TGGAACACCACCATGAAAAGGGCTCAAGTGCTGTGGCTTC-ACCAACTATACGGGATTT 596
           | ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    541  GTTGGAACACCACCATGAAAAGGGCTCAAGTGCTGTGGCTTCNNCCAACTATACGGGATTTT 600
Qy    597  GAGGACTTCACCCTACTTCAAAGAACAGTGCCTTCCCCCA-TTCTGTTGCAAT--GAC   653
           || ||| |||||||||||||||||| || ||||||||||||| |||||||||||  ||||
Db    601  GAAGANTCA-CCTACTTCAAAGAAAAGTGCCTTTCCCCCATTTCTGTTGCAATTGACA  659
Qy    654  AACGTCACCAACACAGCCAATGAAA 678
           |||||||||||||||||||||||||
Db    660  AACGTCCCCAACACAGCCAATTGAA 684
```

RESULT 13
V61147
ID   V61147 standard; cDNA; 729 BP.
AC   V61147;
DT   06-JAN-1999 (first entry)
DE   5' cDNA sequence of prostate tumour clone N1-1862.
KW   Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS   Homo sapiens.
PN   WO9837093-A2.
PD   27-AUG-1998.
PF   25-FEB-1998; U03492.
PR   09-FEB-1998; US-020956.
PR   25-FEB-1997; US-806099.
PR   01-AUG-1997; US-904804.
PA   (CORI-) CORIXA CORP.
PI   Dillon DC, Xu J;
DR   WPI; 98-609886/51.
PT   Polypeptides comprising immunogenic portions of prostate proteins -
PT   used in a vaccine for the treatment of prostate cancer
PS   Claim 3; Page 40; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 729 BP; 157 A; 203 C; 165 G; 191 T;

```
Query Match         43.5%;  Score 560.8;  DB 1;  Length 729;
Best Local Similarity  94.9%;  Pred. No. 3.3e-157;
Matches  650;  Conservative  0;  Mismatches  26;  Indels  9;  Gaps  7;
```

```
Qy    1    AGCCAGGCGTCCCTGCCTGCCACTCAGTGCAACACCCGGAGCTGTTTGTCCTTT  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2    AGCCAGGGCGTCCCTGCCTGCCACTCAGTGCAACACCCGGAGCTGTTTGTCCTTT 61

Qy    61   GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62   GTGGANCCTCAGCAGTNCCCTCTTCAGAACTCANTGCCAAGANCCCTGAACAGGAGCCA  121

Qy    121  CCATGCAGTTGCTTCAGCTTCATTAAGACCATGATCATCCTCTTCAATTGCTCATCTTTC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122  CCATGCAGTTGCTTCAGCTTCATTAAGACCATGATCATCCTCTTCAATTGCTCATCTTTC 181
```

```
PS   Claim 3; Page 39; 130pp; English.
CC   The present sequence is a new DNA which encodes an immunogenic portion
CC   of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC   can be used as a vaccine for the treatment of prostate cancer. The DNA
CC   was identified by analysis of a subtracted cDNA library obtained by
CC   subtracting a prostate tumour cDNA expression library with a normal
CC   tissue cDNA library.
SQ   Sequence 751 BP;      188 A;      196 C;      167 G;      156 T;

     Query Match              41.0%;   Score 528.4;  DB 1;  Length 751;
     Best Local Similarity    88.0%;   Pred. No. 1.5e-147;
     Matches  654;  Conservative  0;  Mismatches  75;  Indels  14;  Gaps  9;

Qy   61 GTGGAGCCTCAGCAGTTCCCTCTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA 120
         |||-|-|||  ||||-||  |||  ||| -| -|| |-  ||||||||||||||
Db  741 GTGGGGNCTTCNGAGTTCNTTNNTTNAGAAGTCA--GGCCNNGGNCCTGAACAGGAGCCA 684

Qy  121 CCATGCAGTGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTC 180
         |-|||| |||| -||| -|| ||||-||||||  ||||  - ||||  - ||||
Db  683 CNANGCAG-GGTTCANGTTNATNAAGACCANGNAGATCTTGTNCAATNGGNTCATCGTTN 625

Qy  181 TGTGT---GGTGCAGCCCTGTTGGCAGTGGGCATCT-GGGTGTCAATCGATGGGCATCC 236
         |-  -   ||||||||||||||||||||||||| ||| ||||||||| ||||||||||
Db  624 NGNGNGTGGGGCAGCCCCTGTTGGCAGTGGGGTCAATNGATGGGGCATCC 565

Qy  237 TTTCTGAAGATCTTCGGG---CCACTGTCGTCCAGTGCCATGC-AGTTTGTCAACGTGGG 292
         |||-||||||| ||- ||  ||||||||||||| |||||||||| ||||||||||N||||
Db  564 TTTTTGAAGATTTTTNGGGCCACTGTCGGNTCCCAGTGCCATGCCAGTTTGTCAACGNGGG 505

Qy  293 CT-ACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGCTCTTGG-TTTCCTGGGCTGCTAT 350
         |  |||||||||||||||||||||  |||  ||||||| || || |||||  |||||N|||
Db  504 GTNATTTCTTCATCGCAGCCGGCGGNGTNGTGGTCTTTGNTNTTGGTTTCANGGGCTGNTAT 445

Qy  351 GGTGCTAAGACTGAGAGCAAGTGTGCCCCTGGTGGCATCT-GGGTGTCAATCGATGGGCAC 410
         |||| |||||||||||||||||||  ||| ||||||| |||| |||||||| |||||N|||
Db  444 GGTGNTAAGACTGAGAGCAAGTGCCCTGGTGGCGTCCTTGGTGGCAC 385

Qy  411 TTCATTGCTGAGGTTGCAGCTGCGTGTGGCCTTGGTGTACACCACAATGGCTGAGCAC 470
         ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  384 TTCATTGCTGAGGTNGCAGCTGCGTGTGGTGTACCACCACAATGGCTGAGCAC 325

Qy  471 TT-CCTGACGTTGCTGGTAGTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTT 529
         |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  324 TTCCCTGACGTTGCTGGTAGTGCCATCAAGAAAGATTATGGTTCCCAGGAAGACTT 265

Qy  530 CACTCAAGTGTGGAACACCACCACCATGAAAGGCTCAAGTGCTGTGCTTCACCAACTATAC 589
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  264 CACTCAAGTGTGGAACNCTACNCACCATGAAAGGCTCAAGTGCTGTGCTTCACCAACTATAC 205

Qy  590 GGATTTTGAGGACTCACCCAACACAGCCATGAAACCTGCCACCAAGCAAAAGGCTCGTTGCAA 649
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  204 GGATTTNGAGGACTCACCNTACNTCAAAGAGAACCTGCCACCAAGCAAAAGGCTCGTTGCAA 145

Qy  650 TGACAACGTCACCAACACAGCCAATGAAACCTGACAATCCGAACTAATGCAGTCACCGTGGG 709
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  144 TGACAACGTCACCAACACAGCCAATGAAACCTGCCAACAGTCACCGTGGG 85

Qy  710 AGTAGAGGGTTGCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTGGG 769
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   84 AGTAGAGGGTTGCTTCAATCAGCTTTTGTATGACATCCGAACTAATGCAGTCACCGTGGG 25

Qy  770 TGGTGTGGCAGCTGGAATTGGGG 792
         |||||||||||||||||||||||
Db   24 TGGTGTGGCAGCTGGAATTGGGG 2
```

Search completed: September 28, 1999, 12:27:16
Job time: 5256 sec

```
                    GenCore version 4.5
         Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    September 28, 1999, 11:33:02 ; Search time 161.06 seconds
                              (without alignments)
                    733.708 Million cell updates/sec

Title:          US-09-030-606-111
Perfect score:  1289
Sequence:       1 AGCCAGGCGTCCCTCTGCCT..........GTTAAAAAAAAAAAAAAAA 1289

Scoring table:  IDENTITY_NUC

Searched:       176461 seqs, 45836279 residues

Database :      Issued_Patents_NA:*
                1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
                4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
                5: /cgn2_6/ptodata/2/ina/PCTUS9_COMB.seq:*
                6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83.4	6.5	977	4	US-08-855-140-2	Sequence 2, Appli
2	75.4	5.8	1151	4	US-08-807-044-2	Sequence 2, Appli
3	62.8	4.9	1452	4	US-08-807-044-4	Sequence 4, Appli
4	62.8	4.9	1452	5	PCT-US91-04986-1	Sequence 2, Appli
5	56	4.3	687	1	US-08-254-493-2	Sequence 2, Appli
6	56	4.3	1120	1	US-08-254-493-3	Sequence 3, Appli
7	56	4.3	687	3	US-08-408-222B-2	Sequence 2, Appli
8	56	4.3	1120	3	US-08-408-222B-3	Sequence 3, Appli
9	50	3.9	7218	2	US-08-232-463-14	Sequence 14, Appl
10	38.8	3.0	349	1	US-07-603-451A-4	Sequence 4, Appli
11	38.8	3.0	349	1	US-08-060-822A-4	Sequence 4, Appli
12	38.8	3.0	349	5	PCT-US94-05257-4	Sequence 4, Appli
13	36.2	2.8	5975	1	US-08-404-354B-1	Sequence 1, Appli
14	36.2	2.8	5975	1	US-08-314-083B-1	Sequence 1, Appli
15	36.2	2.8	5975	2	US-08-435-675B-1	Sequence 1, Appli
16	36.2	2.8	5975	3	US-08-336-257A-3	Sequence 7, Appli
17	36	2.8	44377	4	US-08-804-227C-7	Sequence 7, Appli
c 18	34	2.6	729	5	PCT-US94-01149-17	Sequence 17, Appl
c 19	33.2	2.6	1943	3	US-08-631-200-11	Sequence 11, Appl
c 20	33.2	2.6	1943	3	US-08-929-553-11	Sequence 11, Appl
c 21	33.2	2.6	1943	4	US-08-922-267A-11	Sequence 11, Appl
c 22	33.2	2.6	1943	4	US-08-936-707A-11	Sequence 11, Appl
c 23	33.2	2.6	1943	4	US-08-936-706A-11	Sequence 11, Appli
c 24	32.6	2.5	912	5	PCT-US94-01149-31	Sequence 31, Appli
25	32.4	2.5	975	2	US-08-671-525B-9	Sequence 9, Appli
26	32.4	2.5	975	2	US-08-572-109B-9	Sequence 9, Appli
27	32.4	2.5	975	3	US-08-842-045-9	Sequence 9, Appli
28	32.4	2.5	975	4	US-08-842-238-9	Sequence 9, Appli
29	32.4	2.5	2574	4	US-08-977-734A-8	Sequence 8, Appli
30	32	2.5	2098	3	US-08-153-848-41	Sequence 41, Appl
31	32	2.5	2098	5	PCT-US93-11153-41	Sequence 41, Appl
c 32	31.8	2.5	1037	1	US-08-462-195-1	Sequence 1, Appli
c 33	31.8	2.5	1037	1	US-08-636-883-1	Sequence 1, Appli
c 34	31.8	2.5	6318	4	US-08-808-793-1	Sequence 1, Appli
c 35	31.8	2.5	501	4	US-08-483-695-28	Sequence 28, Appli
c 36	31.8	2.5	501	4	US-07-965-285-28	Sequence 28, Appl
37	31.8	2.5	1332	4	US-09-057-762-1	Sequence 1, Appli

38	31.2	2.4	2093	1	US-08-287-001A-1	Sequence 1, Appli
c 39	31.2	2.4	1333	4	US-08-288-630-1	Sequence 1, Appli
40	31.2	2.4	2093	5	PCT-US95-09941-1	Sequence 1, Appli
c 41	31	2.4	3707	4	US-08-118-101A-1	Sequence 1, Appli
42	31	2.4	2544	4	US-08-469-412A-6	Sequence 6, Appli
43	30.8	2.4	1054	1	US-08-148-215A-1	Sequence 24, Appl
44	30.6	2.4	2150	1	US-08-299-849B-24	Sequence 25, Appl
45	30.6	2.4	2099	1	US-08-299-849B-25	

```
                              ALIGNMENTS

RESULT 1
US-08-855-140-2
; Sequence 2, Application US/08855140
; Patent No. 5584022
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Bandman, Olga
;   APPLICANT: Goli, Surya K.
;   APPLICANT: Guegler, Karl J.
;   TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Incyte Pharmaceuticals, Inc.
;     STREET: 3174 Porter Drive
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/855,140
;     FILING DATE: Herewith
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Billings, Lucy J.
;     REGISTRATION NUMBER: 36,749
;     REFERENCE/DOCKET NUMBER: PF-0296 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 415-855-0555
;     TELEFAX: 415-845-4166
;     TELEX:
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 977 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: MYOMNOT01
;     CLONE: 779308
US-08-855-140-2

Query Match             6.5%;  Score 83.4;  DB 4;  Length 977;
Best Local Similarity  52.0%;  Pred. No. 6.1e-16;
Matches 248;  Conservative  0;  Mismatches 211;  Indels 18;  Gaps  2;

Qy  121 CCATGCAGTGCTTCAGCTTCATTAGCACCATGATGATCCTTCAATTTGCTCATCTTC 180
            ||||  |||  |  ||| ||| | ||||||||||||||||||| |||||||||| |
Db  126 CCAGGGGCTGCCTCGCTGCTTGAAGTACACATGATGTTCCTCTTCAATTGAATATTCTGGC 185

Qy  181 TGTGTGGTGCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGGCATCCTTTC 240
        |.||||| |  ||| ||| |||  ||||| ||| ||||||||| || |  |||| ||||
```

Query Match   5.8%;   Score 75.4;   DB 4;   Length 1151;
Best Local Similarity   55.9%;   Pred. No. 1.8e-13;
Matches 189;   Conservative   1;   Mismatches 133;   Indels   15;   Gaps   2;

```
Qy  129 TGCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCGTGTGT 188
            ||||| ||  |     | ||  | | |  ||  || | |  | | | | || ||||
Db  226 TGCCTCCAGGCCGTCAAGTACCTCATGTTCGCCCTTCAACCTGCTCNTTCGGCTGGGAGGC 285

Qy  189 GCAGCCCTTGGCAGTGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTGAAGATC 248
            | ||| |||  || || ||||||| |||| | | |           |||| ||  | |
Db  286 TGTGGCGTGCTGGGTGTCGGCATCTGGCT----------GGCCGCCACACAGGGGAGC 333

Qy  249 TTCGGCCACTGTCGTCCAGTGCCATGCAGTTTGTCAACGTGGCTACTTCCTCATCGCA 308
            ||| ||    || ||| :|||||    | | ||    |  ||| |   || ||||||
Db  334 TTCGCCAAGGTGTCCTC---TTCCTTCCCGTCCGTCGGGCTGCCAACCTGCTCATCATC 390

Qy  309 GCCGGCGTTGTGTGTCTTTGCTCTTGGTTTCTCGGGCTGCTATGGTGCTAAGACTGAGAGC 368
            ||||| |  |||  |  ||| |  | |  || ||||| |||| || || |  ||||| | ||
Db  391 ACCGGCGCGCCTTGTCATGGCCATCGGCTTCGTGGGTGCTGCCATCAAGGAGAAC 450

Qy  369 AAGTGTGCCCTGGACGTTCTTCTCATCCTCCCTCCTCCATCTTCATTGCTGAGGTTGCA 428
            ||||| |||| || | || ||||| ||| ||| |  |||| || | ||||||| |||| |
Db  451 AAGTGCCTCCTGCTTCACTTTCTTCCTGCTCTGCTGGTGTCTTCCTGCTGCTGGAGGCCACC 510

Qy  429 GCTGCTGGTGGTCGCCTTGGTGTACACCACAATGGCTGA 466
            || ||  ||  | |||| || |||||| ||| |
Db  511 ATCGCCATCCTCTTCTTCGCCTACACGGACAAGATTGA 548
```

RESULT 3
US-08-807-044-4
; Sequence 4, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Goli, Surya K.
;   TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/807,044
;   FILING DATE: Herewith
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0224 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
;   TELEX:
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1452 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear

```
Db  186 TCTGTGGCTGTGGGCTGCTGGGAGTGGGCATCTGGCTCTCCGTGTCCAAGGCAACTTG 245

Qy  241 TGAAGATCTTCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCC 300
            |  ||        |||         ||||||  |||| ||  | |   |  ||| ||||
Db  246 CCACCTTC---------TCCCCAGCTTCCCTTCGTGTCTGTGTCTGCAGCCAACCTGG 290

Qy  301 TCATCGGCGAGCCGGGTTGTGGTCTTTGCTCTTGGTTTCTCGGGCTGCTATGGTGCTAAGA 360
            ||||| |   |  ||| || || |  ||| |  |  || | ||||  ||| | || |
Db  291 TCATCGCCATAGGCACCATTGTCATGTGACGGGCCTTCGTCGGGTGACCTCGGGGCCATCA 350

Qy  361 CTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCATCCTCCCTCCTCCATCTTCATTGCTG 420
            | ||  ||||||| || || | ||| ||||| |||| | | |  ||||| ||||| |||
Db  351 AGGAAAACAAGTGCCTCCTCCTAGCTTTTCATCGTCGTGTCACCCTCCTAGCAG 410

Qy  421 AGTTGCAGCTGCTGTGTGTCGCCTTGGTGTACA---CCACAATGGCTGAGCACTTCCTGA 477
            || ||||| || ||    ||||  | | |||||   |||| | | ||    || ||  ||
Db  411 AGCTGATCTTACCCATCCTCTTCTTGTCTACATGGACAAGGTGAACGAGACGCCAAGA 470

Qy  478 CGTTGCTGGTAGTGCCTGCCATCAAGAAGAATTATGGTTCCCAGGAGACTTCACTCAAG 537
            ||  | |   |  ||||||  || || | ||| ||  ||    ||  ||  |   |||
Db  471 AGGACCTGAAGGAAGGCCTGCTGCTGTACCACCGAGAACAACGTGGGGGCTGAAGAAACG 530

Qy  538 TGTGGAACACCACCATGAAAGGCTCAAGTGCGTGGCCTTCACCAACTATACGGATT 594
            |||||||| ||| |      |||| ||| ||||  ||     ||| |||| |||| |
Db  531 CCTGGAACATCATCCAGCTGAGGATGCCGATGCGTGGTGTCACTGACTTACACAGACT 587
```

RESULT 2
US-08-807-044-2
; Sequence 2, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Goli, Surya K.
;   TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/807,044
;   FILING DATE: Herewith
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0224 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
;   TELEX:
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1151 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;   LIBRARY: SCORNOT01
;   CLONE: 663655
US-08-807-044-2

; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 180140
US-08-807-044-4

Query Match   4.9%;   Score 62.8;   DB 4;   Length 1452;
Best Local Similarity   51.2%;   Pred. No. 1.4e-09;
Matches 173; Conservative   0; Mismatches 162; Indels   3; Gaps   1;

Qy   257 ACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCCTCATCCGCAGCCGGCGT 316
          ||| ||||||||||| ||| ||| ||| |||||| | || |||| | || |||
Db   196 AGTGTCCTTCCATAACCTCCCCCTCCCCTCACGCTGGGCAATGTGTTTGTCATCGTGGGCTC 255

Qy   317 TGTGGTCTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGC 376
          | | ||||| |||| |||||||||||| || | |||||| |||| |||||||| ||||||
Db   256 TATTATCATGGTAGTTGCCTTCCTGGGCTGCATGGGCTGCTATCATCAAGGAAAAACAAGTGTCT 315

Qy   377 CCTCGTGACGTTCTTCTTCATCCTCCTCCTCCATCTTCATTGCTGAGGTTGCAGCTGCTGT 436
          || || ||||||| |||| || || ||||| | ||||||| || ||||||| |||| |
Db   316 GCTTATGTCGTTCTTCATCCTGCTGCTGATTATCCTCCTTGCTGAGGTGACCTTGGCCAT 375

Qy   437 GGTCGCCTTGGTGTACACCAACAATGCTGAGCACTTCCTGACGTTGCTGGTGAGTGCCTGC 496
          ||| || ||| || | |||| |||||||| ||||| | ||| | |||| |
Db   376 CCTGCTCTTTGTATATGAACAGAAGCTGAATGAGTATGTGGC---TAAGGGTCTGACCGA 432

Qy   497 CATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACCATGAA 556
          || || |||||||| |||| ||| | || |||||||||||||||| |||| |||||||
Db   433 CAGCATCCACCGTTACCACTCAGACAATAGCACAATAGCCAGCGTGGGACTCCATCCAGTC 492

Qy   557 AGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATT 594
          ||| ||||| ||||| ||||| ||| | | |||| |||
Db   493 ATTTCTGCAGTGTGTGGTATAAATGGCACGAGTGATT 530

RESULT 4
PCT-US91-04986-1
; Sequence 1, Application PC/TUS9104986
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Allen, Janet
; APPLICANT: Aruffo, Alejandro
; APPLICANT: Camerini, David
; APPLICANT: Lauffer, Leander
; APPLICANT: Oquendo, Carmen
; APPLICANT: Simmons, David L.
; APPLICANT: Stamenkovic, Ivan
; APPLICANT: Stengelin, Siegfried
; APPLICANT: Amiot, Martine
; TITLE OF INVENTION: Rapid Immunoselection Cloning Method
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee & Associates
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/04986
; FILING DATE: 19910715
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/553,759
; FILING DATE: 13-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/498,809
; FILING DATE: 23-MAR-1990

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/379,076
; FILING DATE: 13-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/160,416
; FILING DATE: 25-FEB-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Wall, Margaret M.
; REGISTRATION NUMBER: 33,462
; REFERENCE/DOCKET NUMBER: 11-88C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-499-8080
; TELEFAX: 303-499-8089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1452 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 74..733
PCT-US91-04986-1

Query Match   4.9%;   Score 62.8;   DB 5;   Length 1452;
Best Local Similarity   51.2%;   Pred. No. 1.4e-09;
Matches 173; Conservative   0; Mismatches 162; Indels   3; Gaps   1;

Qy   257 ACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCCTCATCCGCAGCCGGCGT 316
          ||| ||||||||||| ||| ||| ||| |||||| | || |||| | || |||
Db   196 AGTGTCCTTCCATAACCTCCCCCTCCCCTCACGCTGGGCAATGTGTTTGTCATCGTGGGCTC 255

Qy   317 TGTGGTCTTTGCTCTTGGTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGC 376
          | | ||||| |||| |||||||||||| || | |||||| |||| |||||||| ||||||
Db   256 TATTATCATGGTAGTTGCCTTCCTGGGCTGCATGGGCTGCTATCATCAAGGAAAAACAAGTGTCT 315

Qy   377 CCTCGTGACGTTCTTCTTCATCCTCCTCCTCCATCTTCATTGCTGAGGTTGCAGCTGCTGT 436
          || || ||||||| |||| || || ||||| | ||||||| || ||||||| |||| |
Db   316 GCTTATGTCGTTCTTCATCCTGCTGCTGATTATCCTCCTTGCTGAGGTGACCTTGGCCAT 375

Qy   437 GGTCGCCTTGGTGTACACCAACAATGCTGAGCACTTCCTGACGTTGCTGGTGAGTGCCTGC 496
          ||| || ||| || | |||| |||||||| ||||| | ||| | |||| |
Db   376 CCTGCTCTTTGTATATGAACAGAAGCTGAATGAGTATGTGGC---TAAGGGTCTGACCGA 432

Qy   497 CATCAAGAAAGATTATGGTTCCCAGGAAGACTTCACTCAAGTGTGGAACACCACCATGAA 556
          || || |||||||| |||| ||| | || |||||||||||||||| |||| |||||||
Db   433 CAGCATCCACCGTTACCACTCAGACAATAGCACAATAGCCAGCGTGGGACTCCATCCAGTC 492

Qy   557 AGGGCTCAAGTGCTGTGGCTTCACCAACTATACGGATT 594
          ||| ||||| ||||| ||||| ||| | | |||| |||
Db   493 ATTTCTGCAGTGTGTGGTATAAATGGCACGAGTGATT 530

RESULT 5
US-08-254-493-2
; Sequence 2, Application US/08254493
; Patent No. 5439886
; GENERAL INFORMATION:
; APPLICANT: IKEYAMA, SHUICHI
; APPLICANT: KOYAMA, MASARU
; APPLICANT: MIYAKE, MASAYUKI
; APPLICANT: SENOO, MASAHARU
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
; TITLE OF INVENTION: PRODUCTION THEREOF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US

```
       ZIP:  02109
  COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/08/254,493
       FILING DATE:
       CLASSIFICATION:  435
  PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  US 07/865552
       FILING DATE:  09-APR-1992
  PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  JP 079996-1991
       FILING DATE:  12-APR-1991
  PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  JP 085396-1991
       FILING DATE:  17-APR-1991
  PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  JP 022321-1992
       FILING DATE:  07-FEB-1992
  ATTORNEY/AGENT INFORMATION:
       NAME:  RESNICK, DAVID S.
       REGISTRATION NUMBER:  34235
       REFERENCE/DOCKET NUMBER:  41777
  TELECOMMUNICATION INFORMATION:
       TELEPHONE:  (617) 523-3400
       TELEFAX:  (617) 523-6440
       TELEX:  200291 STRE UR
  INFORMATION FOR SEQ ID NO:  2:
       SEQUENCE CHARACTERISTICS:
       LENGTH:  687 base pairs
       TYPE:  nucleic acid
       STRANDEDNESS:  double
       TOPOLOGY:  linear
       MOLECULE TYPE:  cDNA to genomic RNA
       HYPOTHETICAL:  NO
       ANTI-SENSE:  NO
       ORIGINAL SOURCE:
       ORGANISM:  HUMAN
       CELL TYPE:  BREAST CARCINOMA
       CELL LINE:  ZR-75-1
  US-08-254-493-2
```

```
Query Match             4.3%;       Score 56;     DB 1;    Length 687;
Best Local Similarity  49.0%;       Pred. No. 1.1e-07;
  Matches  149; Conservative    0; Mismatches  155;    Indels     0;    Gaps      0;

Qy   130 GCTTCAGCTTCATTAAGACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGTGGTG 189
             || || ||    ||  ||   || ||| || ||| || ||  ||
Db    17 GCACCAAGTGCATCAAATACCTGCTGTTCGGATTAACTTCATCTTCTGGCTTGCCGGGA 76

Qy   190 CAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTCTGAAGATCT 249
             | ||| |   |||  || || |  |||| |  ||||
Db    77 TTGCTGTCTCCTTGCCATTGGACTATGGCTCCGATTCGATGGGGTGGCTACTTCCTCAG 136

Qy   250 TCGGGCCACTGTCGTCCAGTGCCATGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAG 309
                                     || || || |   |||| || || ||| |
Db   137 AGCAAGAAACTAATAATAATTCCAGCTTCTACACAGGAGCTATATTCTGATCGGAG 196

Qy   310 CCGGGCCGTTGGTCTTGCTCTGCGTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCA 369
             ||||| |  |  |  ||| |||||||||| || ||
Db   197 CCGGCGCCCTCATGATGCTGCGGCTTCCTGGGCGTTCGCGGGGTGTGCAGGAGTCCC 256

Qy   370 AGTGTGCCCCTCGTGACGTTCTTCTTCATCCTCCTCCATCTTCATTGCCTGAGGTTGCAG 429
             | | |||  ||| ||||  || | |   ||| |||
Db   257 AGTGCATGCTGGGACTGTTCTTCGGCTTCCTTCTCTTTGGTGATATTCCGCCATTGAAATAGCTG 316

Qy   430 CTGC 433
             |||
Db   317 CGGC 320
```

```
RESULT 6
US-08-254-493-3
  ; Sequence 3, Application US/0825493
  ; Patent No. 5439886
  GENERAL INFORMATION:
  ; APPLICANT:  IKEYAMA, SHUICHI
  ; APPLICANT:  KOYAMA, MASARU
  ; APPLICANT:  MIYAKE, MASAYUKI
  ; APPLICANT:  SENOO, MASAHARU
  ; TITLE OF INVENTION:  MONOCLONAL ANTIBODY, POLYPEPTIDES AND
  ; TITLE OF INVENTION:  PRODUCTION THEREOF
  ; NUMBER OF SEQUENCES:  7
  ; CORRESPONDENCE ADDRESS:
  ; ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN
  ; STREET:  130 WATER STREET
  ; CITY:  BOSTON
  ; STATE:  MASSACHUSETTS
  ; COUNTRY:  US
  ; ZIP:  02109
  ; COMPUTER READABLE FORM:
  ; MEDIUM TYPE:  Floppy disk
  ; COMPUTER:  IBM PC compatible
  ; OPERATING SYSTEM:  PC-DOS/MS-DOS
  ; SOFTWARE:  PatentIn Release #1.0, Version #1.25
  ; CURRENT APPLICATION DATA:
  ; APPLICATION NUMBER:  US/08/254,493
  ; FILING DATE:
  ; CLASSIFICATION:  435
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER:  US 07/865552
  ; FILING DATE:  09-APR-1992
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER:  JP 079996-1991
  ; FILING DATE:  12-APR-1991
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER:  JP 085396-1991
  ; FILING DATE:  17-APR-1991
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER:  JP 022321-1992
  ; FILING DATE:  07-FEB-1992
  ; ATTORNEY/AGENT INFORMATION:
  ; NAME:  RESNICK, DAVID S.
  ; REGISTRATION NUMBER:  34235
  ; REFERENCE/DOCKET NUMBER:  41777
  ; TELECOMMUNICATION INFORMATION:
  ; TELEPHONE:  (617) 523-3400
  ; TELEFAX:  (617) 523-6440
  ; TELEX:  200291 STRE UR
  INFORMATION FOR SEQ ID NO:  3:
  ; SEQUENCE CHARACTERISTICS:
  ; LENGTH:  1120 base pairs
  ; TYPE:  nucleic acid
  ; STRANDEDNESS:  double
  ; TOPOLOGY:  linear
  ; MOLECULE TYPE:  cDNA to genomic RNA
  ; HYPOTHETICAL:  NO
  ; ANTI-SENSE:  NO
  ; ORIGINAL SOURCE:
  ; ORGANISM:  HUMAN
  ; CELL TYPE:  BREAST CARCINOMA
  ; CELL LINE:  ZR-75-1
  ; FEATURE:
  ; NAME/KEY:  CDS
  ; LOCATION:  112..795
  ; FEATURE:
  ; NAME/KEY:  mat_peptide
  ; LOCATION:  115..795
  US-08-254-493-3
```

```
         Query Match          4.3%;     Score 56;    DB 1;    Length 1120;
```

```
Best Local Similarity   49.0%;      Pred. No. 1.4e-07;
Matches 149; Conservative    0; Mismatches 155;      Indels    0;  Gaps   0;

Qy  130 GCTTCAGCTTCATTAAGACCACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGTGGTG 189
            ||  ||| | ||| | | || ||  |  ||  || |  |||| |||      |
Db  128 GCACCAAGTGCATCAAATACCTGCTGTTCGGATTTAACTTCATCTTCTGGCTTGCCGGGA 187

Qy  190 CAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTGAAGATCT 249
           || | ||| | | |  | ||  |
Db  188 TTGCTGTCTTCCTTGCCATTGGACTATGGCCTCCGATTCGACCTCCAGACCAAGAGCATCTTCG 247

Qy  250 TCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAG 309
            ||  |  |  || ||| || ||||
Db  248 AGCAAGAAACTAATAATAATTCCAGCTTCACACAGGAGTCTATATTCTGATCGGAG 307

Qy  310 CCGGGGTTGTGGGTCTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCA 369
            |  | | || || |||||||||||||
Db  308 CCGGCGCCCTCATGATGCTGGTGGGCCTTCCTGGGCTGCTGCGGGAGTCCC 367

Qy  370 AGTGTGCCCTGCGTGACGTTCTTCTTCATCCTCCCTCATCTTCATTGCTGAGGTTGCAG 429
           | | | || || ||||| ||
Db  368 AGTGCATGCTGGGACTGCTTCTTCGGCTTCTTCCGCTTCCTCTTGGTGATATTCGCCATTGAAATAGCTG 427

Qy  430 CTGC 433
            | ||
Db  428 CGGC 431
```

RESULT 7
US-08-408-222B-2
; Sequence 2, Application US/08408222B
; Patent No. 5776727
; GENERAL INFORMATION:
; APPLICANT: Ikeyama, Shuichi
; APPLICANT: Koyama, Masaru
; APPLICANT: Miyake, Masauyki
; APPLICANT: Senoo, Masaharu
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,222B
; FILING DATE: 22-MAR-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,493
; FILING DATE: 06-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-079996-1991
; FILING DATE: 12-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-085396-1991
; FILING DATE: 14-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP-022321-1992
; FILING DATE: 07-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S.
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 41777-DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400

; TELEFAX: 617-523-6440
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to genomic RNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: human
; CELL TYPE: breast carcinoma
; CELL LINE: ZR-75-1
US-08-408-222B-2

```
Query Match  4.3%;     Score 56;    DB 3;   Length 687;
Best Local Similarity   49.0%;      Pred. No. 1.1e-07;
Matches 149; Conservative    0; Mismatches 155;      Indels    0;  Gaps   0;

Qy  130 GCTTCAGCTTCATTAAGACCACCATGATGATCCTCTTCAATTGCTCATCTTTCTGTGTGGTG 189
            ||  ||| | ||| | | || ||  |  ||  || |  |||| |||      |
Db   17 GCACCAAGTGCATCAAATACCTGCTGTTCGGATTTAACTTCATCTTCTGGCTTGCCGGGA 76

Qy  190 CAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGGCATCCTTTCTGAAGATCT 249
           || | ||| | | |  | ||  |
Db   77 TTGCTGTCTTCCTTGCCATTGGACTATGGCCTCCGATTCGACCTCCAGACCAAGAGCATCTTCG 136

Qy  250 TCGGGCCACTGTCGTCCAGTGCCATGCAGTTGTCAACGTGGGCTACTTCCTCATCGCAG 309
            ||  |  |  || ||| || ||||
Db  137 AGCAAGAAACTAATAATAATTCCAGCTTCACACAGGAGTCTATATTCTGATCGGAG 196

Qy  310 CCGGGGTTGTGGGTCTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCA 369
            |  | | || || |||||||||||||
Db  197 CCGGCGCCCTCATGATGCTGGTGGGCCTTCCTGGGCTGCTGCGGGAGTCCC 256

Qy  370 AGTGTGCCCTGCGTGACGTTCTTCTTCATCCTCCCTCATCTTCATTGCTGAGGTTGCAG 429
           | | | || || ||||| ||
Db  257 AGTGCATGCTGGGACTGCTTCTTCGGCTTCTTCCGCTTCCTCTTGGTGATATTCGCCATTGAAATAGCTG 316

Qy  430 CTGC 433
            | ||
Db  317 CGGC 320
```

RESULT 8
US-08-408-222B-3
; Sequence 3, Application US/08408222B
; Patent No. 5776727
; GENERAL INFORMATION:
; APPLICANT: Ikeyama, Shuichi
; APPLICANT: Koyama, Masaru
; APPLICANT: Miyake, Masauyki
; APPLICANT: Senoo, Masaharu
; TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dike, Bronstein, Roberts & Cushman
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/408,222B
; FILING DATE: 22-MAR-1995

Patent No. 5670367
GENERAL INFORMATION:
    APPLICANT: DORNER, F.
    APPLICANT: SCHEIFLINGER, F.
    APPLICANT: FALKNER, F. G.
    TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
    NUMBER OF SEQUENCES: 52
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Foley & Lardner
        STREET: 1800 Diagonal Road, Suite 500
        CITY: Alexandria
        STATE: VA
        COUNTRY: USA
        ZIP: 22313-0299
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/232,463
        FILING DATE:
        CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US/07/935,313
        FILING DATE:
    APPLICATION NUMBER: EP 91 114 300.6
        FILING DATE: 26-AUG-1991
    ATTORNEY/AGENT INFORMATION:
        NAME: BENT, Stephen A.
        REGISTRATION NUMBER: 29,768
        REFERENCE/DOCKET NUMBER: 30472/114 IMMU
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (703)836-9300
        TELEFAX: (703)683-4109
        TELEX: 899149
    INFORMATION FOR SEQ ID NO: 14:
        SEQUENCE CHARACTERISTICS:
            LENGTH: 7218 base pairs
            TYPE: nucleic acid
            STRANDEDNESS: single
            TOPOLOGY: linear
        IMMEDIATE SOURCE:
            CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match   3.9%; Score 50; DB 2; Length 7218;
Best Local Similarity   6.2%; Pred. No. 2.8e-05;
Matches   26; Conservative 217; Mismatches 177; Indels   0; Gaps   0;

Qy   26 CTCAGTGGCAACACCCGGAGCTGTTTGTCCTTTGTCCTTTGTGGAGCCTCAGCAGTTCCCCTCTT 85
           || ||| ||||||| ||||||| :::::::YYYYYYYYYYYYYYYYYYYYYYYY
Db 1038 CTTGGCTGCAGGTGCAGGGAGCTTGGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1097

Qy   86 CAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCACCAGTGCTTCAGCTTCATTAA 145
        ::::::YYYYYYYYYYYYYYYYYYYYY :: : :::YYYYYYYYYYYYYYYYYYYYYY ::
Db 1098 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1157

Qy 146 GACCATGATGATCCTCTTCAATTGCTCAATTTCTCTGTGTGCAGCCCTGTTGGCAGT 205
        :   :::::::::::::::YYYYYYY :: :::::::YYYYYYYYYYYYYY :::
Db 1158 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1217

Qy 206 GGGCATCTGGTGTCAATCGATGGGGCATCCTTCTGAAGATCTTCGGGCCACTGTCGTC 265
        ::: ::: YYYYYYYYYYYYYY ::::::::::YYYYYYYYYYYYYYYYYYYYYYYY
Db 1218 YYYYYYYYYYYYY :: :::::YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1277

Qy 266 CAGTGCCATGCAGTTTGTCAACGTGGCTACTTCCTCATCGCAGCCGCGGCCGTTGTGGTCTT 325
        :: :::::::::YYYYYYYYYYYYYYYYYYYYYYYY :::: :
Db 1278 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1337

Qy 326 TGCTCTTGGTTCCTCGGCGTGCTATGGTGCTAAGACTGAGAGACAAGTGCCCTCGTGAC 385
        : : ::::: :: : :::: ::

    CLASSIFICATION: 536
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: 08/254,493
        FILING DATE: 06-JUN-1994
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: JP-079996-1991
        FILING DATE: 12-APR-1991
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: JP-085396-1991
        FILING DATE: 14-APR-1991
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: JP-022321-1992
        FILING DATE: 07-FEB-1994
    ATTORNEY/AGENT INFORMATION:
        NAME: Resnick, David S.
        REGISTRATION NUMBER: 34,235
        REFERENCE/DOCKET NUMBER: 41777-DIV
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 617-523-3400
        TELEFAX: 617-523-6440
    INFORMATION FOR SEQ ID NO: 3:
        SEQUENCE CHARACTERISTICS:
            LENGTH: 1120 base pairs
            TYPE: nucleic acid
            STRANDEDNESS: double
            TOPOLOGY: linear
        MOLECULE TYPE: cDNA to genomic RNA
        HYPOTHETICAL: NO
        ANTI-SENSE: NO
        FRAGMENT TYPE:
        ORIGINAL SOURCE:
            ORGANISM: human
            CELL TYPE: breast carcinoma
            CELL LINE: ZR-75-1
        FEATURE:
            NAME/KEY: Coding Sequence
            LOCATION: 112...795
            OTHER INFORMATION: E Mat peptide
US-408-222B-3

Query Match   4.3%; Score 56; DB 3; Length 1120;
Best Local Similarity   49.0%; Pred. No. 1.4e-07;
Matches   149; Conservative   0; Mismatches 155; Indels   0; Gaps   0;

Qy 130 GCTTCAGCTTCATTAAGACCATAAGATGATCCTCTCAATTGCTCATCTTTCTGTGTGGTG 189
            ||| |:| ||| |||||||| |||||||| |:| |||| ::::::::: :: ||
Db 128 GCACCAAGTGCATCAAATACCTGCTGTTCGGATTTAACTTCATCTTCTGGCTTGCCGGGA 187

Qy 190 CAGCCCTGTTGCCAGTGGGCATCGGGTGTCAATCGATGGGGCATCCTTTCTGAAGATCT 249
        || | |:| ||||| || ::|:||| |:|:|:|||||| ||:|| | | |
Db 188 TTGCTGTCCTTGCCATTGGACTATGGCTCCGATTCGACTCTCAGACCACAGAGCATCTTCG 247

Qy 250 TCGGGCCACTGTCGTCGTCGCCAGTGCCATGCAGTTTGTCAACGTGGCTACTTCCTCATCGCAG 309
          | | |: |||||| |||:|:| :|| | ||:|:||| |:| |||||| | |||
Db 248 AGCAAGAAACTAATAATAATAATTCCAGCTTCTCACCAGGAGTCTATATTCTGATCGGAG 307

Qy 310 CCGGGCTTGTGGTCTTTGCTCTTGGTTCCTGGCGTGCTATGGTGCTAAGACTGAGAGCA 369
        |||||| | | | |||||||||||| || | ||| || | | || ||| |:|:||
Db 308 CCGGGCCCCTCATGATGCTGGTGGGCTTCCTGGGCTGCTTGCGGGGCTGTGCCAGGAGTCCC 367

Qy 370 AGTGTGCCCCTCGTGACGTTCTTCATCCTCCTCAATCTTCATTGCTGAGGTTGCAG 429
        |||| | ||||:| |||| :|||| || |||| || | ||| || || |
Db 368 AGTGCATGCTGGGACTGTTCTTCGGCTTCCGGCTTCCGGCTTCCTTCTTGGTGATATTCCCCATTGAAATAGCTG 427

Qy 430 CTGC 433
        || |
Db 428 CGGC 431

RESULT   9
US-08-232-463-14
; Sequence 14, Application US/08233463

Db 1338 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1397

Qy   386 GTTCTTCTTCATCCTCCTCCTCATCTTCATTGCTGAGGTGCAGCTGCTGTGGTCGCCTT 445
         :::::: :::::: :: || || :: || || ||
Db 1398 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCCTATCTCTTT 1457

RESULT 10
US-07-603-451A-4
; Sequence 4, Application US/07603451A
; Patent No. 5202420
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Diamond, Gill
;   APPLICANT: Zasloff, Michael
;   TITLE OF INVENTION: Antimicrobial Peptides, DNA
;   TITLE OF INVENTION: Sequences and Methods for the Production and Use Thereof
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
;   ADDRESSEE: No. 5202420ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: Pennsylvania
;   COUNTRY: USA
;   ZIP: 19103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: WORDPERFECT 5.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/603,451A
;   FILING DATE: 19901025
;   CLASSIFICATION: 514
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Patricia A. Schreck
;   REGISTRATION NUMBER: 33,777
;   REFERENCE/DOCKET NUMBER: CH-0005
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (215) 568-3100
;   TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 349 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-07-603-451A-4

Query Match   3.0%; Score 38.8; DB 1; Length 349;
Best Local Similarity 52.5%; Pred. No. 0.012;
Matches   85; Conservative   0; Mismatches 77; Indels 0; Gaps 0;

Qy 1128 TGCCCCCTAGGCCTAGTGGTGATCCCAGTGCTCTACTGGGGGATGAGAGAAAGGCATTTT 1187
         || || :: :: :::: || : | |||| :| |||| :| |||| |:||
Db   184 TGGCACCTGTGTTGGGCGGGCAGTAAAATGCTGTGAAAGAAGTAAAAGAGGGCCAAGAC 243

Qy 1188 ATAGGCCTGGGCATAAGTGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAAGGCACTTCA 1247
         | |||| || :: ||||| | :: ||||||:|:||| |||| | |||
Db   244 ACAGCCGGGATCAATGCCCAGTCAGAAACTCGCGCCCTTTGACAGAGCGGTCTAAAATTTAA 303

Qy 1248 AAATGCATAAACCTGTTACATGTTAAAAAAAAAAAAAA 1289
         | |||| :| | ||| |||||||||||||||||
Db   304 ACCAGAATAAATTTTGTTCAAAGTTAAAAAAAAAAAAAAAA 345

RESULT 11
US-08-060-822A-4
; Sequence 4, Application US/08060822A
; Patent No. 5432270
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Diamond, Gill
;   APPLICANT: Zasloff, Michael
;   TITLE OF INVENTION: No. 5432270e1 Antimicrobial Peptides
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5432270ris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: Pennsylvania
;   COUNTRY: USA
;   ZIP: 19103
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: WORDPERFECT 5.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/060,822A
;   FILING DATE: 19930511
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 07/991,200
;   FILING DATE: December 15, 1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Rebecca Ralph Gaumond
;   REGISTRATION NUMBER: 35,152
;   REFERENCE/DOCKET NUMBER: CH-0408
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (215) 568-3100
;   TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 349 base pairs
;   TYPE: nucle4c acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
US-08-060-822A-4

Query Match   3.0%; Score 38.8; DB 1; Length 349;
Best Local Similarity 52.5%; Pred. No. 0.012;
Matches   85; Conservative   0; Mismatches 77; Indels 0; Gaps 0;

Qy 1128 TGCCCCCTAGGCCTAGTGGTGATCCCAGTGCTCTACTGGGGGATGAGAGAAAGGCATTTT 1187
         || || :: :: :::: || : | |||| :| |||| :| |||| |:||
Db   184 TGGCACCTGTGTTGGGCGGGCAGTAAAATGCTGTGAAAGAAGTAAAAGAGGGCCAAGAC 243

Qy 1188 ATAGGCCTGGGCATAAGTGAAATCAGCAGAGCCTCTGGGTGGATGTGTAGAAGGCACTTCA 1247
         | |||| || :: ||||| | :: ||||||:|:||| |||| | |||
Db   244 ACAGCCGGGATCAATGCCCAGTCAGAAACTCGCGCCCTTTGACAGAGCGGTCTAAAATTTAA 303

Qy 1248 AAATGCATAAACCTGTTACATGTTAAAAAAAAAAAAAA 1289
         | |||| :| | ||| |||||||||||||||||
Db   304 ACCAGAATAAATTTTGTTCAAAGTTAAAAAAAAAAAAAAAA 345

RESULT 12
PCT-US94-05257-4
; Sequence 4, Application PC/TUS9405257
; GENERAL INFORMATION:
;   APPLICANT: Bevins, Charles L.
;   APPLICANT: Diamond, Gill
;   APPLICANT: Zasloff, Michael
;   TITLE OF INVENTION: Novel Antimicrobial Peptides
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & Norris
;   STREET: One Liberty Place - 46th Floor
;   CITY: Philadelphia
;   STATE: Pennsylvania
;   COUNTRY: USA
;   ZIP: 19103

COMPUTER READABLE FORM:
  MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
  COMPUTER: IBM PS/2
  OPERATING SYSTEM: PC-DOS
  SOFTWARE: WORDPERFECT 5.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: PCT/US94/05257
  FILING DATE: 11-MAY-1994
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/060,822
  FILING DATE: May 11, 1993
  APPLICATION NUMBER: 07/991,200
  FILING DATE: December 15, 1992
ATTORNEY/AGENT INFORMATION:
  NAME: Rebecca Ralph Gaumond
  REGISTRATION NUMBER: 35,152
  REFERENCE/DOCKET NUMBER: CH-0408
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (215) 568-3100
  TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 349 base pairs
    TYPE: nucle4c acid
    STRANDEDNESS: double
    TOPOLOGY: linear

PCT-US94-05257-4

Query Match   3.0%; Score 38.8; DB 5; Length 349;
Best Local Similarity 52.5%; Pred. No. 0.012; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 77;

Qy 1128 TGCCCCCTAGGCCTAGTGGTGATCCCAGTGCTCTACTGGGGATGAGAGAAAGGCATTTT 1187
Db   184 TGGCACCTGTGTTGGGCGGCAGTAAAATGCTGTAGAAGAAGAGTAAAGAAGGCCAAGAC 243

Qy 1188 ATAGCCTGGGCATAAGTGAAATCAGCAGAGCCTCTGGGTGGGATGTGTAGAAGGCACTTCA 1247
Db   244 ACAGCCGGGATCAATGCCCAGTCAGAAACTGCGCCCTTTGACAGAGCGTCTAAAATTTAA 303

Qy 1248 AAATGCATAAACCTGTTACAATGTTAAAAAAAAAAAAAAAAA 1289
Db   304 ACCAGAATAAATTTGTTCAAAGTTAAAAAAAAAAAAAAA 345

RESULT 13
US-08-404-354B-1
; Sequence 1, Application US/08404354B
; Patent No. 5618720
; GENERAL INFORMATION:
;   APPLICANT: Ellis, Steven Bradley
;   APPLICANT: Williams, Mark E.
;   APPLICANT: Harpold, Michael Miller
;   APPLICANT: Schwartz, Arnold
;   APPLICANT: Brenner, Robert
;   TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brown, Martin, Haller & McClain
;   STREET: 1660 Union Street
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92101-2926
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/404,354B

FILING DATE: 15-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/314,083
  FILING DATE: 28-SEP-1994
  APPLICATION NUMBER: US 07/914,231
  FILING DATE: 13-JUL-1992
  APPLICATION NUMBER: US 07/603,751
  FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
  NAME: Seidman, Stephanie L
  REGISTRATION NUMBER: 33,779
  REFERENCE/DOCKET NUMBER: 6362-53192
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 619-238-0999
  TELEFAX: 619-238-0062
  TELEX:
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 5975 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
  FEATURE:
    NAME/KEY: Coding Sequence
    LOCATION: 79..5700
  OTHER INFORMATION:
US-08-404-354B-1

Query Match   2.8%; Score 36.2; DB 1; Length 5975;
Best Local Similarity 53.1%; Pred. No. 0.4; Indels 68; Gaps 0;
Matches 77; Conservative 0; Mismatches 0;

Qy 274 TGCAGTTTGTCAACGTGGGCTACTTCCTCATCGACCAGCGGCGTTGTGGTCTTTGCTCTTG 333
Db 344 TGGAGTACTTCTTCCTCACCGTCTTCCATCGAAGCCCATGAAGATCATCGCCTACG   403

Qy 334 GTTTCCTGGGCTGTCTATGGTGTCTAAGACTGAGAGACTGCCCTCGTGACGTTCTTCT   393
Db 404 GCTTCCTGTTCCACCAGGACGCCTACCTGCGCGAGCGGCTGGAACGTGCTGGACTTCATCA 463

Qy 394 TCATCCTCCTCCATCTTCATTGC 418
Db 464 TCGTCTTCCTGGGGGTCTTCACGGC 488

RESULT 14
US-08-314-083B-1
; Sequence 1, Application US/08314083B
; Patent No. 5686241
; GENERAL INFORMATION:
;   APPLICANT: Ellis, Steven Bradley
;   APPLICANT: Williams, Mark E.
;   APPLICANT: Harpold, Michael Miller
;   APPLICANT: Schwartz, Arnold
;   APPLICANT: Brenner, Robert
;   TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brown, Martin, Haller & McClain
;   STREET: 1660 Union Street
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92101-2926
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/314,083B
FILING DATE: 28-SEPT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53191
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-08-314-083B-1

Query Match   2.8%;   Score 36.2; DB 2; Length 5975;
Best Local Similarity   53.1%;   Pred. No. 0.4;
Matches   77; Conservative   0; Mismatches   68; Indels   0; Gaps   0;

Qy 274 TGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGCTCTTG 333
Db 344 TGGAGTACTTCTTCCTCACCGTCTTCCATCGGAGCCCGCCATGAAGATCATCGCCTACG 403

Qy 334 GTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCT 393
Db 404 GCTTCCTGTTCCACCAGGACGCCTACCTGCGCAGCGGCTGGAACGTGCTGGACTTCATCA 463

Qy 394 TCATCCTCCTCCTCCATCTTCATTGC 418
Db 464 TCGTCTTCCTGGGGTCTTCACGGC 488

RESULT 15
US-08-435-675B-1
; Sequence 1, Application US/08435675B
; Patent No. 5710250
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Arnold
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/435,675B
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,083
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53193
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-08-435-675B-1

Query Match   2.8%;   Score 36.2; DB 2; Length 5975;
Best Local Similarity   53.1%;   Pred. No. 0.4;
Matches   77; Conservative   0; Mismatches   68; Indels   0; Gaps   0;

Qy 274 TGCAGTTTGTCAACGTGGGCTACTTCCTCATCGCAGCCGGCGTTGTGGTCTTTGCTCTTG 333
Db 344 TGGAGTACTTCTTCCTCACCGTCTTCCATCGGAGCCCGCCATGAAGATCATCGCCTACG 403

Qy 334 GTTTCCTGGGCTGCTATGGTGCTAAGACTGAGAGCAAGTGTGCCCTCGTGACGTTCTTCT 393
Db 404 GCTTCCTGTTCCACCAGGACGCCTACCTGCGCAGCGGCTGGAACGTGCTGGACTTCATCA 463

Qy 394 TCATCCTCCTCCTCCATCTTCATTGC 418
Db 464 TCGTCTTCCTGGGGTCTTCACGGC 488

Search completed: September 28, 1999, 11:33:05
Job time: 2065 sec

Wed Sep 29 14:27:02 1999